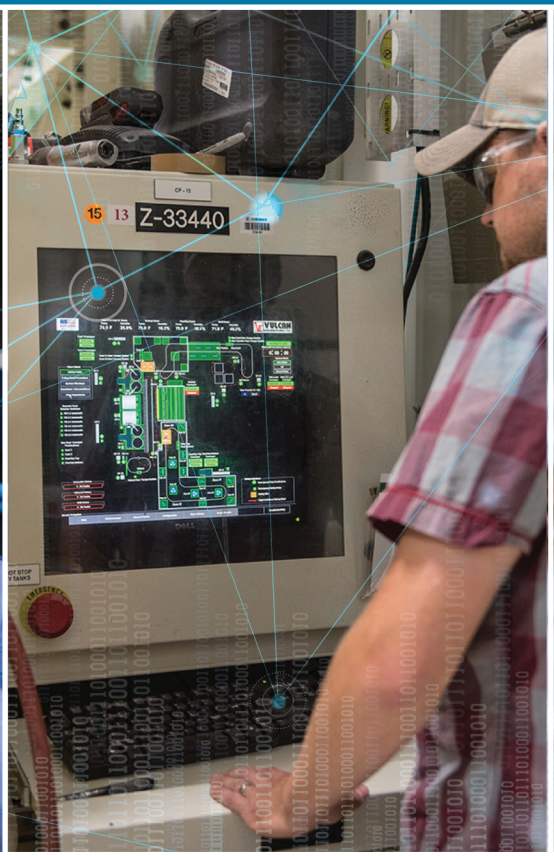
Artificial Intelligence and Advanced Analytics in Indiana:

An Initial Discussion of Industry Needs and University Capabilities



Prepared by: TEConomy Partners LLC
Prepared for: BioCrossroads
January 2020

BioCrossroads®

TECONOMY PARTNERS LLC

Contents

BioCrossroads

Prologue

Technology is a part of our daily lives changing the way we access information, communicate, and interact with each other and the world. It has found its way into our homes, schools, and businesses. And, new and sophisticated technologies have changed how healthcare is delivered from the bench to the bedside—in how discoveries are made to how our doctors select and manage care. The impact has also been felt in Indiana's manufacturing and insurance industries. At the heart of these changes are data, how it is collected, mined and analyzed to drive discovery, enable simulations, forecasting and other complex processing. Data analytics, artificial intelligence (AI), and the Internet of Things are terms that are most often used in describing how technology is applied to innovation and work. Of course, the most vital part of the equation, but often overlooked, is the human element- from those who design and advance technology to those whose work has changed due to its application as more automated settings become the norm.

Data analytics and artificial intelligence are making a mark on Indiana's innovation economy, which includes life sciences, advanced manufacturing, agriculture, information technology and energy systems. The life sciences sector alone is an important economic driver for the State of Indiana providing one in every ten Hoosier jobs, 20 per cent of all tax revenues, and several billions of dollars in annual R&D expenditures by our life sciences industrial partners. These numbers grow when we include the impact of healthcare delivery with more than 150,000 employed by Indiana's hospitals and outpatient facilities.

Given the speed of change, BioCrossroads set out to better understand the readiness of the State of Indiana to participate in how innovation is changing our economy. Having core competencies and access to talent to understand and deploy advanced analytics across advanced industry applications will be especially crucial to state and regional economic futures. BioCrossroads decided to undertake this study to provide an assessment of the current status of the region in advanced data sciences and to ask central questions. Can we grow or attract the necessary talent? Do we have sufficient academic research and industry/academe collaborative engagement to keep innovation happening here?

The report begins with an overview of advanced analytics and the digitally-enabled economy. The next sections are the result of data analysis along with information from a series of interviews conducted by TEConomy with major employers in Indiana as well as leadership at Indiana University, Purdue University, and the University of Notre Dame. These provided an opportunity to understand industry needs and opinions regarding AI and advanced analytics in Indiana as well as understand our university research capabilities and core competencies, including a view of talent development and entrepreneurship. The final section provides a set of recommendations for the state within the context of national and global trends.

The findings are significant. Technological change is impacting our economy at a pace that is unprecedented. Significant investments have already been made—with many more to come—by both our corporate community and each of our major research universities to build capability and attract talent. While our university investments occur in Indiana, and indeed allow for the attraction of talent and capital from outside the state and our federal government in the form of grants, our corporate community has been making significant investments in these capabilities outside of Indiana and will continue to do so if not presented with the relevant technological capability and capacity here in Indiana. In order to ensure that more of that investment is committed to Indiana, and to ensure that the most interesting work that provides meaningful employment continues to happen here, we must take a more intentional

approach to promote engagement between and among our corporations and universities. In doing so, we have the opportunity to attract, grow and retain more of the talent that is trained by our excellent research universities in those skills that are in the highest demand, ensure that our existing workforce is provided opportunities for upskilling with new and expanded continuing education programs, and anchor our corporate community in order to drive Indiana's economy and improve our quality of life.

This is an important and timely report. And certainly, it is appropriate here to thank those whose efforts have made it possible: the Lilly Endowment, through a generous grant to the CICP Foundation on behalf of BioCrossroads, that provided the essential funding; the many members of the life sciences community, those driving our innovation economy in manufacturing and insurance, and our leaders at each of our major research universities, who contributed information and participated in interviews; my colleague Nora Doherty, who led this project for BioCrossroads; and our consultants at TEConomy Partners, who know both Indiana and the innovation sector well, and have drawn on their substantial expertise to provide a helpful, comprehensive and urgently needed study.

We encourage the reader to thoughtfully consider the information, analyses and recommendations presented here and actively engage in advancing Indiana's technology driven economy.

Sincerely,

Patricia A. Martin
President and CEO, BioCrossroads
January 2020

BioCrossroads
300 North Meridian Street, Suite 950
Indianapolis, IN 46204

By:
TEConomy Partners, LLC
Simon Tripp, *Principal and Senior Director*
Ryan Helwig, *Principal and Project Director*
Joe Simkins, *Senior Economist*

BioCrossroads
Patricia Martin, *President and CEO*
Nora Doherty, *EVP*
Vanessa Barth, *CSO*
Brian Stemme, *SVP, Engagement*

With thanks to:

Central Indiana Corporate Partnership, Inc.
David Johnson, President and CEO
Nathan Ringham, Project Manager
Beth Bechdol, AgriNovus IN
Mark Howell, Conexus IN
Jason Kloth, Ascend IN
Michael Langellier, Techpoint
Paul Mitchell, ESN

AIT Bioscience
Shelby Anderson, CSO

Anthem
Gail Boudreaux, President & CEO
Rajeev Rananki, Chief Digital Officer

Community Health Network
Ram Yeleti, EVP, Chief Physician Executive
Patrick McGill, Chief Analytics Officer
Michael Shrift, CMO

Cook Regentec
Rob Lyles, President
Brad King, Director Strategic Development

Corteva Agriscience
Steve Webb, R&D Director

Covance
John Spegele, VP

Cummins
Tom Linebarger, CEO
Sherry Aaholm, CIO
David Leach, Exec Dir, Data & Advanced Analytics

Elanco
Aaron Schacht, EVP

Eli Lilly and Company
David Ricks, Chairman & CEO
Vipin Gopal, VP, Chief Data & Analytics Officer

Indiana University
Michael McRobbie, President
Fred Cate, VP for Research
Brad Wheeler, VP IT & CIO
Anantha Shekhar, Assoc VP Research & Clinical Affairs

IU Health
Dennis Murphy, President & CEO
Mark Lantzy, SVP & CIO (former)
Seung Park, SVP & CHIO

OneAmerica
J. Scott Davison, Chairman, President & CEO
Gene Berry, CIO
Andrew Michie, SVP

Purdue University
Mitch Daniels, President
Jay Akridge, Provost & EVP
Dan Hirleman, Chief Corporate & Global Partnerships Officer
Theresa Mayer, EVP Research & Partnerships
Sunil Prabhakar, Director of IDSI
Patrick Wolfe, Dean, College of Sciences

Roche Diagnostics
Jack Phillips, CEO (former)
Ketan Paranjape, VP Diagnostics Information Solutions

RollsRoyce Corporation
Andrew Baker, Government Relations
Kevin Frasier, Leader IT Innovation
Paul Jeray, Data Scientist

Subaru of Indiana Automotive
Bob Wu, VP Info Systems

University of Notre Dame
Rev. John Jenkins, President
Tom Burish, Provost
Bob Bernhardt, VP for Research

Cover photo credit:
OrthoWorx
ADVANCING ORTHOPEDICS for generations to come.

Supported by a grant from:
Lilly Endowment Inc.

Executive Summary

Advancements in science and technology provide fuel for forces of change. They bring opportunities for development of new and improved products, but also have the power to disrupt existing enterprise and business models. Indiana, with an economy rooted in manufacturing and clusters of advanced industries, is particularly exposed to these forces – forces that present not only opportunities for economic growth but also threats to businesses and industries that fail to adapt or recognize changes on the horizon.

The recent "Clusters and Disruptors" report, published by the Central Indiana Corporate Partnership (CICP)[1] highlighted a broad series of technologies that present disruption threats and opportunities for Indiana's manufacturers, life sciences industries, digital enterprises, and others (Figure ES-1).

Figure ES-1: Three Domains of Disruptive Technology, and Associated Individual Disruptive Technologies

Biological	Physical	Cyber/Digital
1. Next-Gen Sequencing 2. Gene Editing 3. Synthetic Biology 4. Regenerative Medicine and Tissue Engineering 5. Metabolic Engineering 6. Bionanotech and Nanomedicine 7. Cyber-Biological Systems/ Implantables	8. Additive Manufacturing 9. Advanced Robotics 10. Autonomous Vehicles 11. Energy Storage 12. Alternative and Renewable Energy 13. Advanced Materials 14. Nanotechnology 15. Electric Vehicles 16. Alternative Mass Transit	17. Artificial Intelligence 18. Virtual and Augmented Reality 19. Cybersecurity 20. Cloud 21. Big Data Analytics and Associative Intelligence 22. Natural Language Processing 23. Quantum Computing 24. Information Validation 25. Edge Computing 26. Mobile Internet 27. The Internet of Things 28. Blockchain Technology

Advanced Analytics – A Cross-cutting Enabler of Change

Advancements in analytical and data sciences, up to and including Artificial Intelligence (AI), are proving to be cross-cutting enablers of advancements in other disruptive technologies. The rapid growth of computational horsepower, low-cost data storage and high-speed data transfer, in combination with fundamental and applied innovations in machine learning, AI and associated algorithms is empowering advancements in the ability to manipulate large scale datasets, often in near real time. This, in turn, is enabling the automation of processes and decisions – processes and decisions that previously were part of a worker's job description. This is being recognized for its potential to significantly impact our economic future and the future characteristics of work.

1 *Simon Tripp and Ryan Helwig. "CLUSTERS & DISRUPTORS: Envisioning Central Indiana's Economic Future in a Time of Change." Produced by TEConomy Partners, LLC. for the Central Indiana Corporate Partnership. September 2018*

For state, regional and local economic developers, a primary focus has always been the development and protection of jobs – particularly jobs paying robust, family-sustaining levels of pay. Overall, advanced and disruptive technologies have tended to create as many, or more, jobs than they destroy. Historically, new technologies tend towards increasing productivity and enhancing GDP growth, which in turn supports new job generation. But, at a regional or local level, the potential for negative economic impacts is far more real and employment growth impacts from industry disruptions less than assured.

A major concern for labor markets is that the convergence of technologies, in combination with rapid advancements in AI and advanced analytics, will likely enable the emergence of intelligent automation across a much broader suite of jobs than previously anticipated.

The smaller a region is in spatial scale and population the less likely it is to have the human capital skills, university R&D specializations, specific educational skills, and other competitive advantages required to respond to new job skill demands and industry needs in fast emerging sectors. The late 19th Century and almost the entire 20th Century provided opportunities, rooted in technological change and innovation, that brought about a rising economic tide that lifted a broad swath of different regional economies. **In recent years, however, it is becoming evident that the modern technological economy is stratifying regions across a place-based win/lose continuum** with high performers seeing an agglomeration of self-reinforcing economic advantages in terms of industry clusters, human capital and financial capital, and poor performers struggling to gain or retain meaningful momentum in fast growing new and expanding industries.

A Concern for Indiana

Having core competencies and access to talent able to understand and deploy advanced analytics across advanced industry applications will be especially crucial to business fortunes moving forward – and thus will be critically important to state and regional economic futures. Within Indiana, many regions across the state are characterized by clusters of manufacturers and associated advanced industries that power the local economy today but are also vulnerable to the forces of change empowered by advanced analytics and associated automation.

Gartner Research has created a working definition of the concept of "advanced analytics":

"Advanced Analytics is the autonomous or semi-autonomous examination of data or content using sophisticated techniques and tools, typically beyond those of traditional business intelligence (BI), to discover deeper insights, make predictions, or generate recommendations. Advanced analytic techniques include those such as data/text mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing, neural networks."

For the purpose of this study, TEConomy considers the techniques and methodologies included across a broad continuum of technological sophistication to be included in the definition of advanced analytics, but with a primary focus on the use of analytics tools and platforms within **applied, industry-facing contexts** to drive process automation, advanced insights, or other value-added tasks in products and services being advanced by companies.

This study has been carried out by TEConomy in close consultation with BioCrossroads—and BioCrossroads is an appropriate principal sponsor for this work, since so much of Indiana's leading life sciences sector is already so substantially driven and disrupted by these strategic technologies. But as TEConomy observed in our 2018 Clusters and Disruptors report, the implications of these rapid and converging technology developments for Indiana's economy range well beyond life sciences and healthcare, and in fact extend to all of Indiana's advanced industry sectors that fall within the various branded industry initiatives of CICP. Additionally, a parallel set of intensive discussions by CICP leadership involving a wide range of major Indiana industry CEOs—in which chief executives expressed serious concerns regarding Indiana's readiness to meet their specific needs for technology-intensive R&D and appropriate talent—has only brought further urgency to the need for this report.

Accordingly, and against this background, BioCrossroads—in coordination with CICP leadership—has determined to lead this effort by TEConomy to provide an assessment of the current status of the region in advanced data sciences and to determine where there are strengths, weaknesses, opportunities and threats related to data analytics that are of relevance to regional economic futures. Central questions given that core competencies in advanced data analytics are increasingly essential drivers of regional competitiveness: Can we grow or attract the necessary talent (undergraduate and graduate)? Do we have sufficient academic research and industry/academe collaborative engagement to keep innovation happening here?

In terms of changing the current profile of employment and work across the U.S., the impact of advanced data analytics will primarily be felt through "automation." Every job contains a series of tasks that combine into the worker's occupation – and the universe of tasks across occupations varies across a continuum from "routine and easily automated" to "complex/creative and hard to automate".[2] The world of work is thus being changed in varying levels by advanced data analytics and associated automation. Recent analysis by Mark Muro and colleagues at Brookings finds that:

- "Approximately 25 percent of U.S. employment (36 million jobs in 2016) will face high exposure to automation in the coming decades (with greater than 70 percent of current task content at risk of substitution). At the same time, some 36 percent of U.S. employment (52 million jobs in 2016) will experience medium exposure to automation by 2030, while another 39 percent (57 million jobs) will experience low exposure."[3]

A major concern for labor markets, including in Indiana, is that **the convergence of multiple technologies, in combination with AI and advanced analytics, will likely enable the emergence of intelligent automation across a much broader suite of jobs than previously anticipated.** Hence the need to understand and quantify the current situation in the state.

Indiana Talent Output in Advanced Analytics

Advancements in advanced analytics and AI, fields that enable automation, are inherently driven by the innovations of skilled human talent. Educated personnel are required who:

- Understand the fundamentals of data sciences and advanced analytical techniques,
- Understand the application of analytical tools and techniques to specific data sciences problems,
- Understand how to translate data sciences innovations and advancements into domain specific applications (across a diverse range of business sectors and business functions).

Digital literacy is fundamentally important to an individual's ability to thrive in the modern economy. Digital technology and data pervade modern economic and societal activity and are at the core of most expanding job markets. A key sub-component of this is **Data Analytics** – providing capacity to understand, process, manage and use sets of digital information. Building digital literacy is a requirement now within the K-12 education system, but for the purposes of this study (where meeting the needs of Indiana industry in the near-term is critical) the output of Indiana's higher education institutions is a more pressing question.

As a topline finding it is readily apparent that Indiana's research universities have recognized the imperative to produce value-added skill sets in this space, to the extent that data analytics is now seen at Indiana University, Purdue University, and the University of Notre Dame as foundational to a well-rounded university education and being integrated into core university curricula that crosses traditional college and departmental boundaries.

2 *For a detailed discussion of occupations, occupational tasks and the comparative potential impact of automation and AI upon them, see: Mark Muro, Robert Maxim and Jacob Whiton. "Automation and Artificial Intelligence: How Machines are Affecting People and Places." Brookings Institution, Metropolitan Policy Program. January 2019.*

3 *Ibid*

It is also evident that IU, Purdue and Notre Dame have a considerable volume of students currently enrolled in courses of study in advanced data analytics. There are a broad variety of data sciences and advanced analytics programs available in Indiana and more in development (on campus, online, full or part-time). TEConomy's analysis of National Center for Education Statistics (NCES) Integrated Postsecondary Education Data System (IPEDS) data indicate that **3,926 graduates in programs relevant to data sciences, advanced analytics and AI graduated from the three Indiana research universities in 2017.** This will be increasing in the future, given the significant commitment to data analytics programs of study at the universities.

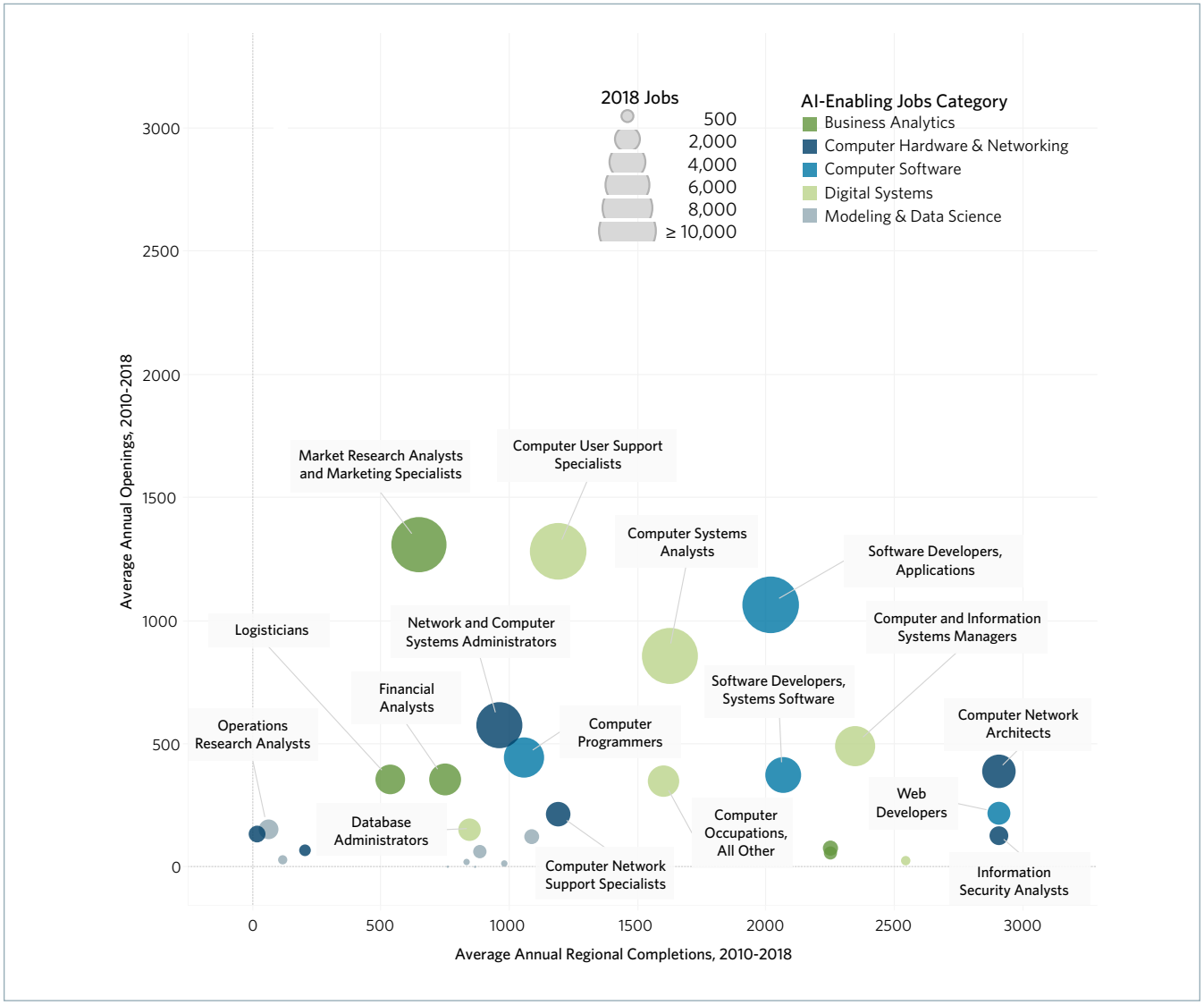
TEConomy's analysis indicates that the state's research universities are generating significant volumes of graduates that are aligned with the types of occupations required to advance AI-enabled applications for industry. It is important, however, to understand the balance between the state's labor force demand in advanced analytics and associated positions and the projected supply of skilled talent from the universities.

Figure ES-2 uses IPEDS and EMSI data on openings and completions (which crosswalks degree types to occupational categories that require highly similar technical skills and knowledge), to compare the "supply" of degree completions from the universities to the "demand" from industry openings over the 2010 to 2018 period in the AI-enabling occupation categories. While these data do not account for precise skills matching or quality of graduates, they do provide general insight into potential shortages or surpluses of talent.

For most occupations that are analytics/AI-enabling or enabled, Indiana's average annual volume of degree completions for 2010 through 2018 exceeds the number of average annual openings in occupations in Indiana for which those degrees are highly aligned.[4] While this finding indicates that Indiana's higher education institutions demonstrate a capacity to educate and train a supply of talent that can satisfy industry openings in AI enabling occupations, meeting this rapidly rising demand would require a level of retention of these graduates in excess of what is currently occurring or could reasonably be expected under current circumstances.

4 *One notable exception in occupations that have high employment footprints in the state includes market research analysts. For this occupation, the number of openings from 2010 to 2018 were significantly higher than the number of degree completions, indicating a potential shortage in talent generation from higher education institutions.*

Figure ES-2: Supply-Demand Balance for AI-Enabling Occupations in Indiana, 2010-2018.



Note: occupations with >1,000 Indiana employees in 2018 annotated
Source: EMSI Occupational Data, EMSI 2019.4; NCES IPEDS Data

These data indicate that there is evidence of a surplus of degree completions that outpaces openings, indicating a pipeline available for industry to leverage. However, feedback from industry stakeholders makes it clear that they experience significant challenges accessing local analytics and AI talent. This difference in narratives indicates that despite evidence of a significant higher education pipeline, there is still a mismatch in the labor market in terms of connecting with this capacity. There are several factors at play and potential explanations:

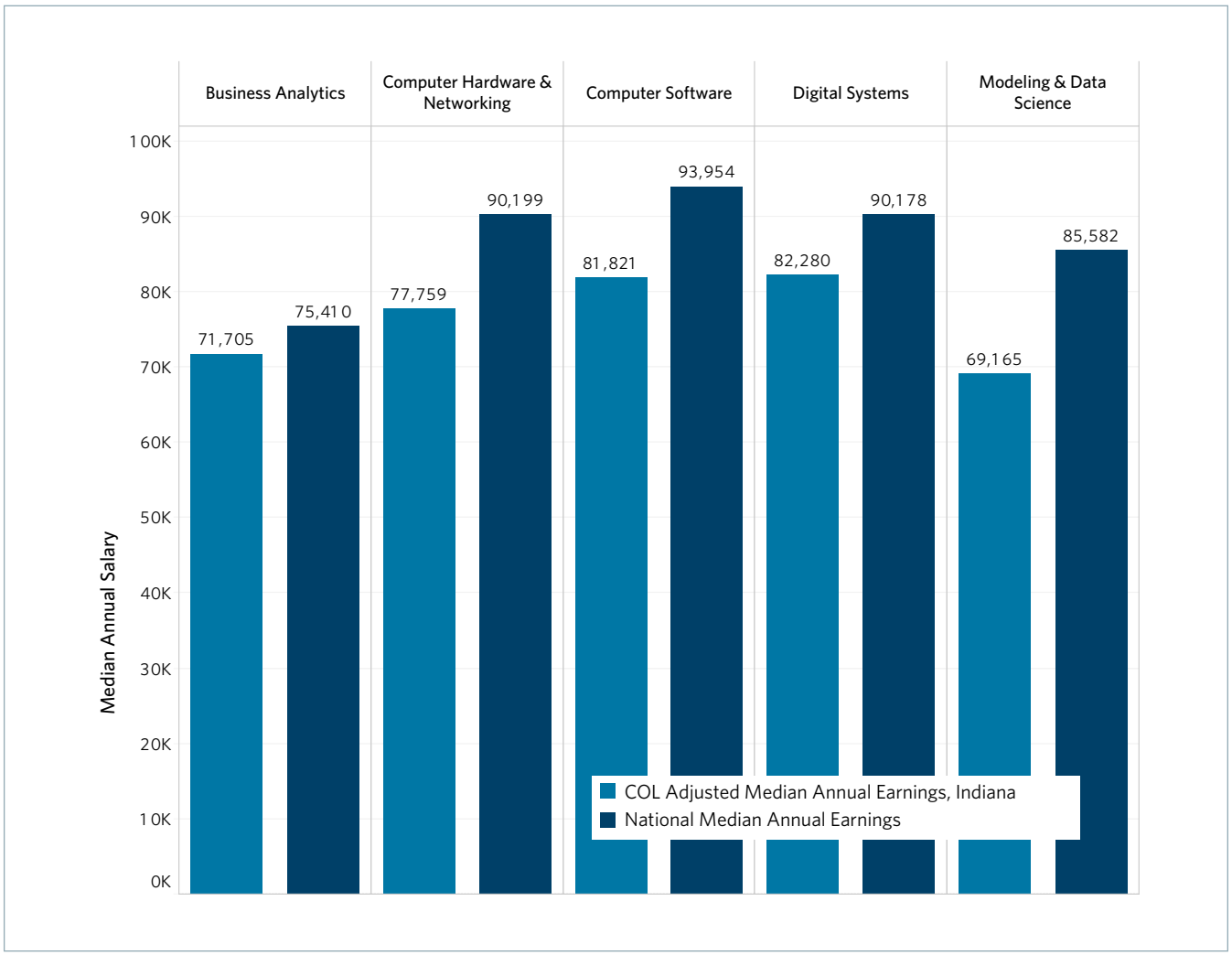
- Much of the graduate output is in areas that are supportive of analytics implementation (computer science, digital systems, etc.) while output in the core areas of "modeling and data sciences" is more modest.
- A number of the major initiatives and programs at the major research universities in Indiana are relatively new and graduate output will be forthcoming.
- There is an evident gap in communications between universities and industry (in both directions) leading to suboptimal awareness and engagement of respective needs and supply.

- Students graduating from the Indiana institutions are being attracted to jobs out-of-state. This continuing development is seen as being driven by three factors:
  - Out-of-state companies are getting to them early in their academic studies (freshman and sophomore years) and building robust relationships through internships. Indiana companies far less so.
  - Students have low levels of awareness of Indiana-based employers, unable to name more than just a few.
  - Salaries being offered in Indiana are generally lower—at all levels of experience and seniority—than national averages (even adjusting for the low comparative cost of living of Indiana).
- IU, Purdue and Notre Dame attract a significant number of students nationally and internationally as each are recognized as well-placed institutions for training in these related areas. While this represents an opportunity to "import" talent to the State of Indiana, a smaller percentage of out-of-state students choose to remain in Indiana after graduation than in-state students, and more than half of in-state students select employment opportunities outside of the state following graduation.
- Indiana's companies are predominantly deployers of advanced analytics in diverse industries and are not at the innovation core of the field in the way that Google, Facebook, Microsoft, IBM or others are perceived to be.

It is also the case that many companies express a preference to hire analytics personnel who have some relevant work experience and experience with their sector. However, it is imperative to understand just how intense the competition is for such workers nationally and internationally. A 2-3 years' experience requirement is likely a luxury that cannot be afforded by current and near-future labor market conditions (unless exceptional pay levels are provided).

Competitive salary levels are an important factor that determines the employment decisions of skilled talent. For the analytics and analytics-enabled occupational space, this aspect of attracting talent is particularly important because of the highly competitive national labor market for these skill sets as well as increased salary expectations that are set by prominent employers in leading technology innovation ecosystems in Silicon Valley and other major coastal metro areas. **TEConomy's analysis finds that cost of living adjusted salaries for Indiana are comparatively low** (Figure ES-3). The overall national median annual earnings level of $84,772 for relevant occupational segments is 21.4percent higher than the cost of living adjusted Indiana median annual earnings level of $69,850, meaning that while labor cost for talent is lower in Indiana, the difference in actual purchasing power relative to national trends potentially makes the state a less attractive destination for employment in these spaces. The disparity is particularly stark in the core "modeling and data sciences" classification. National salaries are consistently higher than cost of living adjusted salaries for Indiana across key segments and occupations, potentially providing some insight into why industry needs are not being met despite signals of talent pipeline surpluses.

Figure ES-3: Median Annual Salaries for AI-Enabling Occupational Segments in Indiana, 2018



Source: EMSI Occupational Earnings Data, EMSI 2019.4

It is likely that the identified earnings differentials may also place a threat not only on the attraction of talent to state jobs but also on retention of skilled analytics talent within Indiana. Figure ES-4 illustrates that the disparity is evident across the full spectrum of analytics associated positions, with the differential expanding for high paid (and presumably higher skill or experience) positions.

Figure ES-4: Distribution of Median Annual Salaries for Advanced Analytics and AI-Enabling Occupations in Indiana and US, 2018

Source: EMSI Occupational Earnings Data, EMSI 2019.4

A central finding of this project is that in advanced analytics and AI Indiana faces a situation that is characterized less in terms of CONTENT (there is talent being produced and a concerted institutional effort to significantly expand this pipeline, and there is a significant presence of faculty expertise in the space and emerging consultative models being developed by the universities) and far more in terms of ENGAGEMENT. Simply put, Indiana's corporations and universities have stronger engagement outside of the state in terms of both R&D and talent-access than they do with their Indiana partners. Intentional engagement between Indiana's corporations and universities needs to be significantly reinforced.

Currently, most of Indiana's skilled analytics graduates flow to out-of-state employers, and out-of-state companies are aggressively engaging talent pipelines early in their education and embedding their presence on university campuses. This reflects the reality that Indiana's data sciences graduates are in high demand in a competitive national labor market and presents a particularly important challenge to address.

Indiana industry stakeholders have also largely focused their efforts to meet their needs for advanced analytics services and talent outside of Indiana, although some very recent shifts towards significant engagement with in-state university partners has started to occur. Most university engagement with Indiana industry today is at smaller scales and revolves around direct relationships with key centers or faculty rather than broader university capabilities or initiatives, with major university corporate partners also tending to be out-of-state. In addition, many university initiatives seeking to grow AI-enabling talent flows and capabilities are relatively new or only recently being brought up to scale, meaning that awareness on the part of regional industry stakeholders is also an issue to be addressed.

The overarching narrative outlined by the convergence of these factors as well as the state's current industry and research position all highlight a missing element for Indiana in growing the regional footprint and competitive position in AI-related spaces. This need can best be described as the development of a shared community of interests involving corporate and university partners, through intentional coordination and communication initiatives, that realigns talent flows and pushes in-state industry needs to universities in an intensive manner.

Connecting Universities and Industry for Analytics Projects

Beyond pure talent development, the application of advanced analytics and AI knowledge at the universities to potential industry development programs is also important. In discussions with industry a series of "use cases" were identified that could form potential foci for joint engagement between universities and Indiana industry.

The identified use cases offered take multiple forms—some are industry-, product- or technology-specific; others are focused on enhancing core business and operational functions; and still others from manufacturers are focused on digital transformation of production and maintenance processes. Each use case would draw from one or more specific advanced analytics capability and/or technologies as set out in Table ES-1.

Table ES-1: Industry-provided Use Cases for Potential Partnering Opportunities with Research Universities and/or Private Industry Solutions Providers.

Applications Area	Specific Use Cases Identified	Advanced Analytics-related Competencies/ Tech Required/Leveraged
Core Business Functions	**Forecasting Sales:** using advanced analytics to improve forecast accuracy of sales to inform and adjust manufacturing, and to configure supply chain. **Decision Support for Managing an R&D Portfolio/Strategy:** for R&D portfolio management, considering the balance of risk, investments and rate of new product introduction – simulation of forward time-based scenarios for decision support. **Enhancing Customer Service:** measurement, analysis, and prediction of customer experience. What is a customer's actual journey as they interact with the company, what are the drivers of dissatisfaction and delight? Given a specific customer, what experience/journey is likely to achieve an outcome (purchase, retention, promotion). **Gauging Marketing Effectiveness:** retrospective and predictive analysis of marketing activities/spend/campaigns. What's the value to the firm (sales, brand, etc.); how should the firm tune marketing better for individual interactions (end customers, producers, etc.)?	• Predictive analytics • Modeling and simulation

Enhancing IT/ Data Analytics Infrastructure	• **Upgrading, Scaling IT Infrastructure:** scaling up, upgrading IT Infrastructure including hardware, installing new apps.
	• **Back-end IT Development Areas:**
	  • enhancing data quality, workflows, operations management – create partnerships giving students access to data to build algorithms.
	  • assistance with product apps in areas such as oncology clinical decision support workflow.
		• Data warehousing/ storage infrastructure, architecture

Applications Area	Specific Use Cases Identified	Advanced Analytics-related Competencies/ Tech Required/Leveraged

Industry/ Product/ Technology Specific Applications	• **Using NLP Tech to Translate Voice to Text:** utilizing Natural Language Processing technology to capture physician notes; align with an industry partner to identify patient candidates for clinical trials.
	• **Leveraging Genomic Data to Inform Clinical Trials:** partner with a third party to use genomic information from a biorepository to compare against clinical trials information.
	• **Analyzing Social Determinants of Patient Health:** a project to leverage structured data regarding social determinants of care (e.g., obesity, smoking cessation).
	• **Studying Opioid Use:** leveraging health systems data and data from the Indiana Health Information Exchange to understand patterns of opioid use.
	• **Leveraging AI and Machine Learning for biomedical/life sciences applications, including:**
	  ▪ for drug discovery and development
	  ▪ for digital health,
	  ▪ algorithms for personalized medicine.
	• **Balancing Risk for Third Party Insurance Sales:** many insurance products for individual sold through institutions such as companies, schools, hospitals etc. How relevant is the institution's reputational/ risk scores to assist and impact individual underwriting?
	• **Predicting Insurance Claims on Initial Indication:** in insurance claims, can prediction of the severity or complexity at first notice of loss become more accurate?
		• Natural Language Processing
		• Data fusion and Big Data analysis
		• Predictive analytics
		• AI and Machine Learning

Conclusions and Recommendations

Indiana has quite robust assets in advanced analytics to build upon and there is momentum starting to build in terms of a number of recent announcements and initiatives on the part of both industry and university that are designed to directly impact the state's position. Indeed, despite the challenges noted there are a number of factors aligning for the state that can be collectively leveraged to build momentum towards a more competitive position for the state:

- Universities have made a significant commitment to building both cross-cutting analytics competency across their campuses, and specific advanced programs, majors and certificate programs.
- Companies and their leadership are now engaging in serious strategic implementation efforts around data sciences and actively seeking to engage with outside high education partners and other supports.
- There is a small but growing base of specialized entrepreneurial companies who are providing services in this space.

What has been lacking has been intentional engagement through coordination and communication between these respective asset bases. In order to create an advanced analytics and AI-enabling ecosystem that truly meets the power and potential of this technology space for Indiana's advanced industries, Indiana must connect and coordinate its assets to grow a broader community of excellence, which in turn can attract and retain business and talent in growing applications areas. This engagement will need to be activated with intentional action connecting university and corporate talent around real-world problems.

TEConomy concludes that engagements with Indiana universities are required across two dimensions:

- A flexible, matrixed approach that is adaptable to the fact that specific engagements will vary highly depending on industry sector applications areas, technology readiness level of the company, and scale of the collaborative effort due to the diverse nature of Indiana industry stakeholders. Development of broader solutions at the industry sector level are unlikely to meet the specific needs of the majority of companies in that sector due to a high level of variance in technological readiness and business norms (with the exception of healthcare analytics as noted below).
- Establish a hub in Indianapolis that can serve to connect companies and talent from each of Indiana's major research universities (in R&D and in education) whether those resources be in Indianapolis or at other in-state campuses. The hub would also be a potential home to multi-disciplinary teams made up of members from companies and universities alike to work toward applying AI and data analytics to solve problems in manufacturing, healthcare and other areas where Indiana has strength.

Realigning talent flows also requires more active relationship-building on the part of industry. A more coordinated approach to managing relationship can help activate programs to improve efforts to attract and retain talent.

In order to address the identified industry needs and challenges and work towards creating a high-functioning ecosystem in advanced analytics and AI-enabling activities, several specific activities are recommended across four thematic areas. These areas include:

- Coordination of Industry Engagement and Continuing to Attract Top Research Talent at Indiana's Research Universities,
- Addressing Demand for Talent and Continuing Education of Existing Workforce,
- Coordinating Access and Collaborations in University-Industry Partnerships and Talent Access in Indianapolis, and
- Enhancement of Indiana's Profile in AI.

1. Coordination of Industry Engagement and Continuing to Attract Top Research Talent at Indiana's Research Universities

The overarching need for Indiana revolves around intentional engagement where intellectual talent from universities and corporations work and interact in both structured and unstructured ways.

- Support and enable traditional sponsored research by taking a more proactive approach to understand data sciences needs of Indiana companies and then making connections to relevant university capacities. Use engagement offices that remain knowledgeable about the comprehensive profile of ongoing university activities in AI-related technologies and facilitating industry relationships with the right university partners.
- Fund and develop applied projects with industry partners where a question of interest is studied by a team made up of individuals with diverse experiences and expertise from academic and corporate institutions working together in a set location for a fixed period of time. The final product could be a presentation, publication, or spin-out. Cohorts working together on such projects necessarily build longer lasting relationships that build community and allow for future opportunities for engagement and problem solving. University personnel can participate through competitive fellowships or sabbaticals.
- Encourage research universities in the state to further adopt interdisciplinary institute models for advancing data sciences within their programs to help aggregate activity institution-wide. Build methodology expertise with recruitment of new talent in mathematical and computations areas that bridge to health and life sciences, manufacturing, and other areas where Indiana has a concentrated industry presence. Concentrate on identifying and attracting talent with National Academy or Lasker award winners, for example, or those rising stars currently working at recognized centers at Stanford, MIT, Northwestern, or Toronto. Establish competitive fellowships, à la MacArthur Foundation, for existing faculty to enable the pursuit of new research directions that build on expertise in data sciences.
- Look to refine Indiana's data assets and technologies and invest in the creation of "data lakes" that can be used to drive innovation in areas of particular concentration and strength such as healthcare and manufacturing. Use AI tools to concentrate data to drive simulations, analyze the past, optimize the present, predict the future, test hypotheses, and explain inferences.

2. Addressing Demand for Talent and Continuing Education of Existing Workforce

Addressing the significant needs of Indiana's industries for skilled talent in data sciences and analytics forms the foundation for growing the state's AI-related capabilities, and primarily involves retaining more graduates from the significant talent pipeline that exist in-state as well as continuing education of existing workforce.

- Establish funding programs for data sciences environments (DSEs) based on industry partners, live-learn community model to bolster pipeline and build student awareness of Indiana opportunities. An initiative in this area would ideally encourage establishment of these programs at all major research universities which could then be connected with the coordination function described above to engage industry partners. Purdue's Data Mine is an example of a developing world class DSE program that is organized around industry engagement and immersive skills-building in data sciences that can serve as a model for other universities.
- Implement incentive programs to retain skilled analytics talent in-state upon graduation. Options to explore include "signing bonus" awards, grants for ongoing education, tuition reimbursement, state tax rebates, and other funding incentives in targeted programs or skill sets aligned with data sciences needs of industry stakeholders.
- Significantly increase the level of in-state industry marketing and presence at hub locations to engage Indiana talent pipelines. Industry stakeholders can pursue more aggressive recruitment strategies in conjunction with coordination efforts described above.
- Maintain comprehensive information on graduates in relevant program areas who leave the state for their first employment position in order to pursue recruitment for senior roles in the future.

- Implement new and expand existing continuing education programs for workforce at Indiana companies. Expand engagement to include visiting lecturers from companies to provide living lab opportunities for problem solving.

3. Coordinating Access and Collaborations in University-Industry Partnerships and Talent Access in Indianapolis

Indianapolis has a significant presence of academic assets and programs of degree study emphasizing advanced analytics – particularly in, but not limited to, healthcare applications and informatics. The IU/Purdue/IUPUI academic constellation in Indianapolis comprises biomedical and health sciences, informatics, computer science, engineering and associated degree programs of IU and Purdue, with activity centrally concentrated at the IUPUI campus in Indianapolis. With Indianapolis being home to significant major corporations (as well as midsize and entrepreneurial growth companies) and major healthcare institutions, there is a built-in regional demand for engagement between these employers who need to apply advanced analytics and the local presence of the universities which can form hubs for talent supply, and regional access points for statewide university capabilities in research, education, training and consultative services.

The fast growth and evolution of careers requiring analytics competencies and credentials, means that it will be increasingly important for education and ongoing certification and recertification courses to be readily accessible to regional employers. A joint university resource hub in Indianapolis that can serve to interface with companies and provide a robust access point for linking companies to the analytics resources of the universities (in R&D and in education) whether those resources be in Indianapolis or at other in-state campuses, would also ensure robust communication of industry needs and projections for talent and skills demands into the future. The resource hub can make the appropriate connections with companies to Indianapolis-based faculty, or to faculty located at other campuses, and provide a core venue for ongoing discussions of talent development and talent retraining program development.

4. Enhancement of Indiana's Profile in AI

For Indiana to establish itself as a credible leader and destination location for the development and deployment of leading technologies relating to artificial intelligence, Internet of Things and advanced data analytics, it must enhance its profile.

- Establish a signature statewide data sciences conference with a rotating in-state university host and awareness-building branding around state industry opportunities and leading applied research activities. Leading companies should be identified to serve as potential sponsors and participants to build knowledge of in-state employment and collaboration opportunities.
- Leverage 16 Tech, an innovation community, as a multi-institutional hub located in the Indianapolis environment focused around healthcare applications of AI and analytics that leverages significant institutional and industry strengths in central Indiana.
- Incentivize entrepreneurial industry-university partnerships in data sciences by mitigating disincentives for faculty to participate in commercialization activities.
    - Support data sciences spin-out companies founded by faculty-entrepreneur partnerships using competitive grant award model and prioritizing companies developed around data sciences services that can be leveraged by larger industry stakeholders
    - Explore funding and incentives for faculty involved in industry consulting and research, including career advancement credit and flexible schedule structures.
- Coordinate state institutions around a common mission and complementary areas of specialization to enable competition at scale with other leading ecosystems. In particular, encouraging institutions to reach consensus on areas of specialization in applied analytics and AI to avoid duplicative investment and establishing protocols to reduce in-state competition for talent and resources is critical to long-term success.

In Summary

Indiana's economic future is and will continue to be impacted by the role of AI and advanced analytics in innovation. Its corporations and universities have made substantial and strategic investments in expanding, improving and building capabilities. Indiana's ability to compete will depend upon its industrial and academic sectors to drive more collaborative engagement with each other in order to grow and maintain an ecosystem that can attract and retain talent while supporting growth. There is a significant pipeline of advanced analytics talent in the state and it is being further expanded through focused investments and innovative program expansions at the major research universities in the state. Furthermore, the universities are structuring transdisciplinary analytics initiatives and consulting access points to enable faculty and other university research personnel to interface with companies on joint R&D projects and specific challenge engagements.

Because of high competition for analytics talent, and aggressive recruitment of Indiana talent by out-of-state enterprises, it is important that Indiana's advanced industries be proactive in interfacing with universities to build early relationships with analytics student populations. It will also be particularly important for industry to engage in intensive internship engagements with students, often across multiple years. It is likely, however, that beyond increasing opportunities for connectivity, Indiana industry will need to raise the salaries it offers for analytics positions to remain competitive for attraction and retention of talent.

Indiana has a significant opportunity to advance its competitive position in the AI-related technologies and talent space through coordinated efforts to connect and realign stakeholder relationships within its ecosystem. The state's path to success lies in leveraging these technologies towards leading applications in its major state industry clusters rather than in pursuing research enterprises in more fundamental "basic data sciences."

I. Introduction

A. Advanced Data Analytics and Its Impact on the Economy

2018 and 2019 have proven to be breakthrough years for general public awareness of the power and promise of **Artificial Intelligence (AI) and Advanced Data Analytics**. Applications of advanced data analytics, including AI, have entered consumer consciousness and moved to the forefront, with many industries seeking to sustain and advance their competitiveness via the intelligent use of large scale and complex datasets. Almost every industry sees potential for application of **advanced data analytics**, with some applications providing incremental improvements in performance and business insights, and other providing for the development of completely new businesses and business models.

For those concerned with regional economic development, there should be little doubt that core competencies in advanced data analytics, including AI, will have a critical impact on regional economic performance. Core competencies in advanced data analytics—from an R&D, business application, and workforce skills perspective—are an increasingly essential driver of regional competitiveness and will only become more so in the future.

For Indiana, this digitally-enabled, advanced data analytics reality is extremely relevant to many of the region's advanced and traded industries. As discussed in the recent "Clusters and Disruptors" report for CICP[5], industries of importance to Indiana are being reshaped by cyber-digital disruptive technologies and the convergence of digital content and control in almost all industries. Driving these changes are advanced data analytics comprising the use of various analytical computer algorithms and statistical techniques acting upon large-scale sets of structured and unstructured data to derive actionable insight, **or** the use of advanced techniques to assess data inputs in real-time for

Advanced Data Analytics and Artificial Intelligence

While AI has become a focus of attention and is certainly an example that is very much on the frontiers of advanced data analytics, it should be viewed as part of a continuum of analytical techniques and applications ranging from traditional statistical tools (such as descriptive statistics, statistical inference, regression techniques, cluster analyses, etc.) through to advanced applications in machine learning and deep learning neural networks.

It should also be noted that there is not a single definition of AI upon which all agree. The Oxford Dictionary defines AI as: "the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages." That certainly captures part of AI, but the full realm of advanced data analytics is far more diverse and complex.

The term **"advanced data analytics"** is perhaps a suitably broad term to use for describing the broader continuum of knowledge, tools, and talent that can leverage AI-based technologies to provide insights. This assessment of Indiana core competencies and potential applications will primarily concentrate on more emerging, frontier areas of data analytics that can be leveraged to deploy machine learning technologies at scale.

5 *Simon Tripp and Ryan Helwig. "CLUSTERS & DISRUPTORS: Envisioning Central Indiana's Economic Future in a Time of Change." Produced by TEConomy Partners, LLC. for the Central Indiana Corporate Partnership. September 2018*

automated decision making and control. Looking across Indiana's advanced industry clusters it is not hard to see the rising importance of advanced data analytics, up to and including AI, for the future of these industries, including in:

- **Agricultural Biosciences** – with the emergence of digital and precision agriculture, integrating physical sensors and data sciences to optimize agricultural production and inputs use.
- **Aircraft/Aerospace** – with the emergence of unmanned and autonomous systems and applications in areas such as remote sensing data collection and analysis, maintenance analytics, air traffic control, etc.
- **Automotive/Mobility** – with autonomous vehicles as an application of AI, and advanced AI-enabled robotics used in manufacturing.
- **Information Technology** – where advanced analytics are obviously a core component of the sector, but especially relevant within Central Indiana where a substantial portion of the IT sector is engaged in advanced marketing, sales and customer service applications, and embedded within financial and insurance services sectors, that see significant application of advanced data analysis.
- **Life Sciences** – where advanced data analytics are deployed in scientific discovery, including drug discovery, all the way through to clinical decision-making support and as an enabling technology for the emergence of personalized medicine.
- **Logistics** – a sector that has been an early adopter of advanced analysis techniques in predictive modeling and optimization, and will increasingly be impacted by AI, robotics, autonomy and associated advanced technology applications.

As work in AI, machine learning, machine vision and associated fields advances, the effect on society, the economy and individual industries will continue to grow. The implementation of advanced analytics will be front and center in shaping regional economic fortunes, the structure of the world of work, and the lives of workers and those who depend upon their income. Ultimately, the "how" of advanced analytics-based economic development has the highest potential to occur along three primary pathways (Table 1):

Table 1: Pathways to Advanced Analytics and AI Enabled Economic Development

Pathway	Illustrative Examples	Effects
Provision of actionable insights rooted in analysis of data that improve the products, services or business operations of an existing organization or business.	Using advanced analytics of customer transaction data to identify meaningful purchase patterns and identify new opportunities for sales. Advanced analytics of clinical trials data to identify sub-populations with genotypes most likely to respond positively to a new drug therapy.	• Enhanced competitiveness • Expanded markets • Reduced product cycle time.
Enabling automation of actions or controls to reduce costs, improve quality and/or enhance productivity.	AI managed inventory, warehouse picking and packing automation for a distributor of products. Automated systems for analysis of pathology or radiologic images to improve diagnostics throughput.	• Improved cost efficiencies • Enhanced productivity • Quality control and repeatability
Spurring new business development in the commercialization of advanced analytics tools, technologies, algorithms and services.	Development of a contract advanced analytics company providing analytics services to small and midsized manufacturers. Commercialization of an AI engine for use in chemical compound analysis and drug discovery.	• New business and business cluster development

There is significant potential for the deployment of advanced analytics to expand regional employment through improving business competitiveness, identifying new market opportunities, and informing the development of new or enhanced products and services to meet market needs, including products and solutions that disrupt traditional markets. At the

same time, the implementation of enhanced AI-enabled automated processes may change the demands on labor as certain tasks would no longer require human intervention across blue-collar, white-collar and even professional jobs.

The emergence of modern advanced analytics means the world of work will undergo profound changes. New personnel with analytics skills will need to be recruited and existing workers re-skilled to adapt to new processes and business applications. Markets and business models have always been subject to change, but the opportunities for data sciences, and especially AI, to bring about radical paradigm shifts over the next two decades may be as significant as the first round of digital IT deployment in the 1980's and 1990's.

B. "Adapt and Thrive" or "Stagnate and Ossify"

Few jobs are the same, or are performed in the same way, as they were 20 years ago. While this is not a universal truth for all legacy occupations, the ways in which most of us perform our work have been impacted by the digital revolution. Many jobs that may look "traditional" on the surface are still very much changed by significant digital content or back end analytical processes. For example, airline pilots, radiologists and pathologists, delivery truck drivers, financial loan officers, may each find themselves working with or alongside digital controls and analytics-based systems now and into the coming decade. The level of impact from these changes can be conceptualized as something akin to a bell curve with a few job types completely unimpacted by digital technology in developed economies as well as a few that have been rendered fully automated or nearly so (such as automotive body welders). However, the vast majority of jobs fall somewhere in the middle of the bell curve where some part of their daily task content is digitally impacted. **As the capabilities of digital systems become more sophisticated and able to take on more complex tasks, the shape of this distribution of impacts will change.**

In the 2018 Clusters and Disruptors report for Central Indiana[6], TEConomy identified a series of disruptive technologies that Indiana would need to pay attention to in considering the ongoing evolution of its existing advanced industry business clusters and the growth of new industries (Figure 1):

In looking across this list of technologies it is evident that all of them will be enabled by or be implemented through digital technologies with advanced data analytics playing a central role in their development, deployment or control. **As a result, analysis of digital data is likely the most powerful agent of change impacting the modern developed economy.**

Figure 1: Three Domains of Disruptive Technology, and Associated Disruptive Technologies

Biological	Physical	Cyber/Digital
1. Next-Gen Sequencing 2. Gene Editing 3. Synthetic Biology 4. Regenerative Medicine and Tissue Engineering 5. Metabolic Engineering 6. Bionanotech and Nanomedicine 7. Cyber-Biological Systems/ Implantables	8. Additive Manufacturing 9. Advanced Robotics 10. Autonomous Vehicles 11. Energy Storage 12. Alternative and Renewable Energy 13. Advanced Materials 14. Nanotechnology 15. Electric Vehicles 16. Alternative Mass Transit	17. Artificial Intelligence 18. Virtual and Augmented Reality 19. Cybersecurity 20. Cloud 21. Big Data Analytics and Associative Intelligence 22. Natural Language Processing 23. Quantum Computing 24. Information Validation 25. Edge Computing 26. Mobile Internet 27. The Internet of Things 28. Blockchain Technology

6 *Simon Tripp and Ryan Helwig. "CLUSTERS & DISRUPTORS: Envisioning Central Indiana's Economic Future in a Time of Change." Produced by TEConomy Partners, LLC. for the Central Indiana Corporate Partnership. September 2018*

C. Changing Jobs, Remaking Skills, Adapting to Change

In terms of changing the current profile of employment and work across the U.S., the impact of advanced data analytics may primarily be felt through "automation." Every job contains a series of tasks that combine into the worker's occupation – and the universe of tasks across occupations varies across a continuum from "routine and easily automated" to "complex/creative and hard to automate".[7] Picking and packing jobs in warehouses, telephone receptionists, manufacturing production line workers, utility meter readers, for example, may be highly susceptible to seeing all or the major part of their work supplanted by automation while others with a high social touch and interpersonal communications such as caregiving for children, the sick or elderly, are unlikely to be automated. The world of work is thus being changed in varying levels by advanced data analytics and associated automation. Recent analysis by Mark Muro and colleagues at Brookings finds that:

- "Approximately 25 percent of U.S. employment (36 million jobs in 2016) will face high exposure to automation in the coming decades (with greater than 70 percent of current task content at risk of substitution). At the same time, some 36 percent of U.S. employment (52 million jobs in 2016) will experience medium exposure to automation by 2030, while another 39 percent (57 million jobs) will experience low exposure."[8]
- "Routine, predictable physical and cognitive tasks will be the most vulnerable to automation in the coming years"[9]
- Impacts in the near-term will be felt disproportionately within jobs that pay relatively low wages. In general, jobs (and their tasks) associated with below baccalaureate levels of education will see the most displacement. As Brookings notes "better-educated, higher-paid earners for the most part will continue to face lower automation threats based on current task content – though that could change as AI begins to put pressure on some higher-wage non-routine jobs."

It should be noted that Indiana is predicted in the Brookings analysis to have the highest percentage of its employment-weighted task load subject to automation (48.7 percent). New York is predicted to have the lowest at 42.4 percent.

For state and local economic developers, a primary concern has always been the protection and development of jobs – particularly jobs paying robust, family-sustaining levels of pay. On a large spatial scale, at the level of a nation or multi-nation economic bloc for example, revolutionary and even disruptive revolutionary technologies have tended to create as many, or more, jobs than they destroy. **New technologies tend towards increasing productivity and enhancing GDP growth, which in turn supports new job generation. But, at a regional or local level, the potential for negative economic impacts is far more real and employment growth impacts from industry disruptions less than assured.** Silicon Valley, for example, has had a winning formula for growth via innovation (especially participation in revolutionary and disruptive revolutionary innovation) but there have also been very real regional losers such as upstate New York with the failure of major employers such as Kodak and Xerox who were eventually unable to adapt to revolutionary change. Similarly, the smaller a region is in spatial scale the less likely it is to have the human capital skills, university R&D specializations, and specific educational skills, and other competitive advantages required to respond to new job skill demands and industry needs in fast emerging sectors. The late 19th Century and almost the entire 20th Century provided opportunities, rooted in technological change and innovation, that brought about a rising economic tide that lifted a broad swath of different regional economies. In recent years, however, **it is becoming evident that the modern technological economy is stratifying regions across a place-based win/lose continuum with high performers seeing an agglomeration of self-reinforcing economic advantages in terms of industry clusters, human capital and financial capital and poor performers being unable to gain meaningful momentum in fast growing new and expanding industries.** This challenging environment is only being exacerbated by the realities of a global economy in which a state like Indiana is not only in competition for jobs against other U.S. cities and regions, but with fast growing global locations as well.

7 *For a detailed discussion of occupations, occupational tasks and the comparative potential impact of automation and AI upon them, see: Mark Muro, Robert Maxim and Jacob Whiton. "Automation and Artificial Intelligence: How Machines are Affecting People and Places." Brookings Institution, Metropolitan Policy Program. January 2019.*

8 *Ibid*

9 *Ibid*

There is also debate taking place among thought leaders in technology and innovation economics regarding the potential impact on future employment levels due to technological change. As noted above, technological innovations have historically tended to create more net new jobs in the economy than they destroy. Innovation creates a virtuous cycle through generating wealth and in turn GDP growth that is distributed through society creating positive multiplier effects that spur further job creation. Newly generated industries have nearly always tended to have significant labor needs as they scale to meet demand. There are, however, weaknesses becoming evident in this economic paradigm with one of the major indicators being the divergence of growth in wage levels from growth in production and GDP levels.[10] As the wealth gains of productivity increases are captured at a corporate level and distributed at a lower rate to labor, traditional gains in standard-of-living from one generation to the next are less assured. Another adverse effect, identified in the Brookings analysis of the impact of automation across the "IT era" (1980-2016), is evident where it is found that:

> *Overall employment has grown in the IT era nationally, but the middle of the wage continuum has been "hollowed out" – with changes in employment and wages greatest at the high and low ends of the wage distribution.*[11]

A major concern for labor markets is that **the convergence of technologies, in combination with AI and advanced analytics, will likely enable the emergence of intelligent automation across a much broader suite of jobs than previously anticipated.** For this to happen, however, certain conditions (largely characterized by technologies) must be met (Table 2):

Table 2: Required Conditions and Technologies Enabling Advanced Analytics and Automation in Practice

Enabling Condition for Automation Expansion	Current Status
Low cost sensors need to be integrated into devices that perform tasks or into any environments that interface with other machines, digital networks or humans.	Already occurring. Development and roll-out of Internet of Things (IoT) moving quite rapidly.
The sensing environment must be networked to allow high-speed communications between system elements.	Low cost 5G and other wireless platforms facilitating deployment.
Data to make associations must be ubiquitously available in a cloud environment and stored in a cost-effective manner.	Companies already porting their data to major cloud service providers, and restructuring their data to facilitate advanced analysis.
Affordable computational horsepower must be available to process extremely large datasets and accommodate real-time inputs.	High performance computing costs are lowering, and companies are able to lease cycles on high performance systems without having to invest in the systems themselves.
Statistical techniques and computational algorithms need to underlie system analytics.	Substantial basic and applied data sciences advancements are occurring in academic, industry, and institutional settings.
Cyber security technologies must secure data and devices.	Ongoing challenge. Likely to be a restraining factor (especially if quantum computing becomes reality).

As can be seen on Table 2, **most of the enabling factors for advanced analytic- based automation to occur are already in place and experiencing rapid advancements.** Cybersecurity remains a primary constraining factor. Beyond the technological reality of the above, there will also be multiple cases where the socio-political reality of regulatory and

10 *The Economic Policy Institute (EPI) notes that "From 1973 to 2016, net productivity rose 73.7 percent, while the hourly pay of typical workers essentially stagnated—increasing only 12.5 percent over 43 years (after adjusting for inflation). This means that although Americans are working more productively than ever, the fruits of their labors have primarily accrued to those at the top and to corporate profits, especially in recent years." The EPI notes that while productivity and total compensation for labor generally tracked each other from the 1940's through the early 1970's, since then there has been a significant divergence, to the point that since 1973 productivity has grown 5.9 times more than pay. Source: "The Productivity – Pay Gap" Updated October 2017. https://www.epi.org/productivity-pay-gap/ Similar data can be observed in Brookings analysis at: https://www.brookings.edu/research/thirteen-facts-about-wage-growth/*

11 *Mark Muro, Robert Maxim and Jacob Whiton. "Automation and Artificial Intelligence: How Machines are Affecting People and Places." Brookings Institution, Metropolitan Policy Program. January 2019.*

public opinion factors will need to be accounted for. A robotic primary care physician making a diagnosis, for example, not only has to be technologically feasible but it must also be "allowed" to be implemented by human regulators across the healthcare environment.[12]

Looking across the six factors on Table 2 it is not hard to see that conditions are already being largely met to enable a rapid expansion of automated decision-making applications of advanced analytics. Combine this with parallel advancements in mobile robotics, autonomous vehicles, remote sensing, high speed wireless data transfer, and relatively low-cost cloud based high performance computing and data storage, etc. and it is not hard to foresee substantial implications for the labor market. One can argue that there will be jobs created in making all these systems, controlling them and maintaining them, but longer-term the future may see:

- Robots themselves manufacturing and repairing robots,
- AI systems designing automation systems and their own applications,
- AI systems controlling complex networks of connected devices at a speed and level of complexity beyond the capability of even the largest teams of humans.

This pending "AI control of AI", or "automated production of automation", is predicted by McKinsey & Company[13] who note that "there doesn't need to be a controller at the center of such intelligence; appropriate action can emerge as a property of the whole system." As devices and systems communicate with each other, and data builds exponentially, **the ability to have fully autonomous systems that are self-organizing, conversational, dynamic and self-adjusting becomes a reality.**

Today's reality is that some jobs are already being lost to, or significantly changed by, automation, AI and associated machine capabilities. Manufacturing and low-skilled work segments are not the only places where rapid changes are disrupting traditional job roles. Radiological images are already being analyzed by intelligent systems without requiring the intervention of a radiologist or other skilled diagnosticians. Skilled credit bureau, mortgage analyst and financial analyst positions are being overtaken by automated systems (albeit introducing new issues with algorithmic bias and data privacy). The sheer speed of automated decision making systems may also remake industries as consumers are empowered by information devices, by choice, and by real-time satisfaction of needs. **Entire industries may have to shift to large-scale deployment of AI-enabled automation, whether they want to or not, in order for firms to make decisions fast-enough to remain competitive.**

12 An interesting current case is that of the Iowa based company IDx. This entrepreneurial company, founded by a professor of ophthalmology at the University of Iowa, developed an automated system for diagnosing diabetic retinopathy – something that previously had to be performed by a qualified human professional. The automated IDx system was quite rapidly adopted in Europe, but FDA and other regulatory restrictions significantly delayed its use in U.S. healthcare.

13 W. Brian Arthur. "Where is technology taking the economy?" McKinsey Quarterly. October 2017.

D. The Education Imperative

Mathematics and English have long been foundational in our education – rightly seen as essential cross-cutting core competencies that provide the ability to comprehend content in other disciplines and successfully navigate the worlds of work and society. The changing landscape described above may also require adding **Digital literacy** to the existing foundation. Digital technology and data pervade modern economic and societal activity and are at the core of most expanding job markets. A key sub-component of this skill set involves **Data Analytics** – providing capacity to understand, process, manage and use sets of digital information. Indiana's research universities have recognized the imperative to produce value-added skill sets in this space, to the extent that data analytics is now seen at Indiana University, Purdue University, and the University of Notre Dame as foundational to a well-rounded university education and being integrated into core university curricula that crosses traditional college and departmental boundaries.

Recent analysis predicts the growth in human jobs (in developed nations) to occur in fields that are especially skewed towards STEM jobs, creative jobs, and jobs requiring digital skills. Analysis by Brookings[14] examined changes in the digital content of 545 occupations embracing 90 percent of the U.S. workforce since 2001. Among their findings:

- "Digitalization is associated with increased pay for many workers and reduced risk of automation, but it is also helping to "hollow out" job creation and wages by favoring occupations at the high and low ends of the pay scale while disfavoring those in the middle."
- "Digitalization is changing the skills needed to access economic opportunity while creating new race- and gender-based access challenges."
- "Changes have been striking. By 2016, the share of employment in occupations with high digital content … more than tripled, from 4.8 to 23 percent of employment." "Employment in occupations with low digital scores … declined precipitously, from 55.7 to 29.5 percent."

The Brookings authors note that digitalization is a powerful force in terms of expanding the American economy, **but to be fully realized improvements in digital education and training are required to "broaden the high skill talent pipeline and ensure that underrepresented groups can connect to an increasingly digital economy."**

Applications of advanced analytics to work activities across industries will impact such a broad range of jobs throughout the economy that the impact will be felt among personnel across all education levels, ages and other demographic categories. The pace of digitally enabled change and the breadth of advanced analytics adoption across industries will be such that skills required cannot be accommodated solely by new entrants to the workforce (those currently coming through the K-12 and traditional higher education pathways). **It will also be necessary to train and re-skill** many in the incumbent workforce, and indeed, the rate of technological change will likely require personnel to re-skill or upgrade skills with increasing regularity, multiple times over their career-span. This requirement will place a premium on having the educational fundamentals that facilitate a life-long learning mindset and access to multiple modalities of affordable and timely education delivery.

It is also the case that many industries that have the potential to be heavily impacted by advanced analytics may not previously have been participants in heavily digitalized or computationally intensive business models. As such, they may lack a culture that is naturally aligned with advanced analytics development and implementation and have only a nominal base of incumbent personnel with formal analytics education. The current trend towards "convergence" in science, technology and business is forcing traditionally bounded industries to expand their horizons to embrace new digital and analytics skillsets. Convergence refers to a trend whereby:

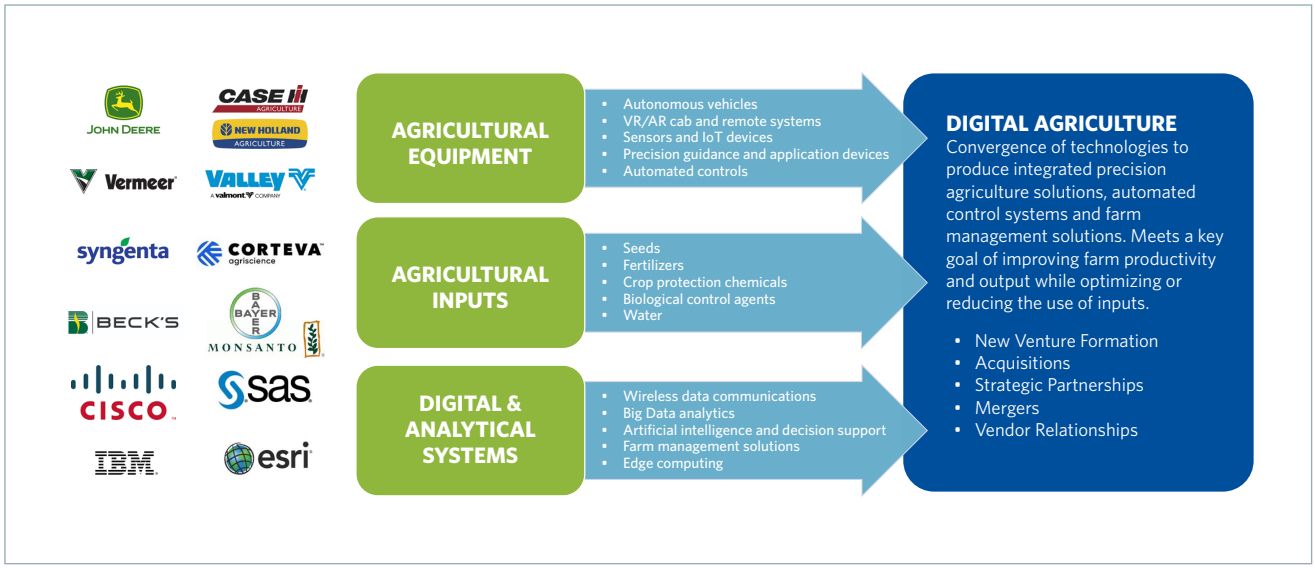
> *Innovation is increasing at the interface between disciplines. Individual discovery is becoming multi-disciplinary and transdisciplinary team discovery. Success comes from integration of multiple cyber, physical and biological*

14 Mark Muro, Sifan Liu, Jacob Whiton and Siddharth Kulkarni. 2017. "Digitalization and the American Workforce". Brookings, Metropolitan Policy Program.

domain elements to derive new products and new solutions to market needs. Companies known for work in one sector, partner with or acquire entities in another to develop multi-component systems and novel combinatorial innovations. Networking, collaboration, adaptability and non-linear thinking come to the forefront of drivers of convergence models.

Advanced analytics represents the most cross-cutting of "disciplines" engaged in and supporting convergence. As an example, in life sciences such convergence is seen in the interface between bioscience and advanced analytics in genomics and bioinformatics, in the interface between agricultural equipment, agricultural inputs and digital technology integration in agriculture (illustrated in Figure 2), and in the performance of certain clinical tasks within healthcare (robotic assisted surgery, machine vision analysis of medical images, AI analysis of electronic medical records, etc.).

Figure 2: Digital/Precision Agriculture as an Example of Industry Convergence and the Roll of Digital and Advanced Analytical Systems



It is now a clear economic reality that advanced analytics enables the development and adoption of disruptive technologies and new business models, together with enabling the refinement and evolution of existing business models, that are required for a region to compete in the global economy. Increasingly, it is a foundational skillset required for personal and corporate success, and an absolute imperative for states such as Indiana to address given the potential impact on core advanced industry and traded industry clusters.

E. Defining "Advanced Data Analytics": From Digital Systems to Digital Intelligence

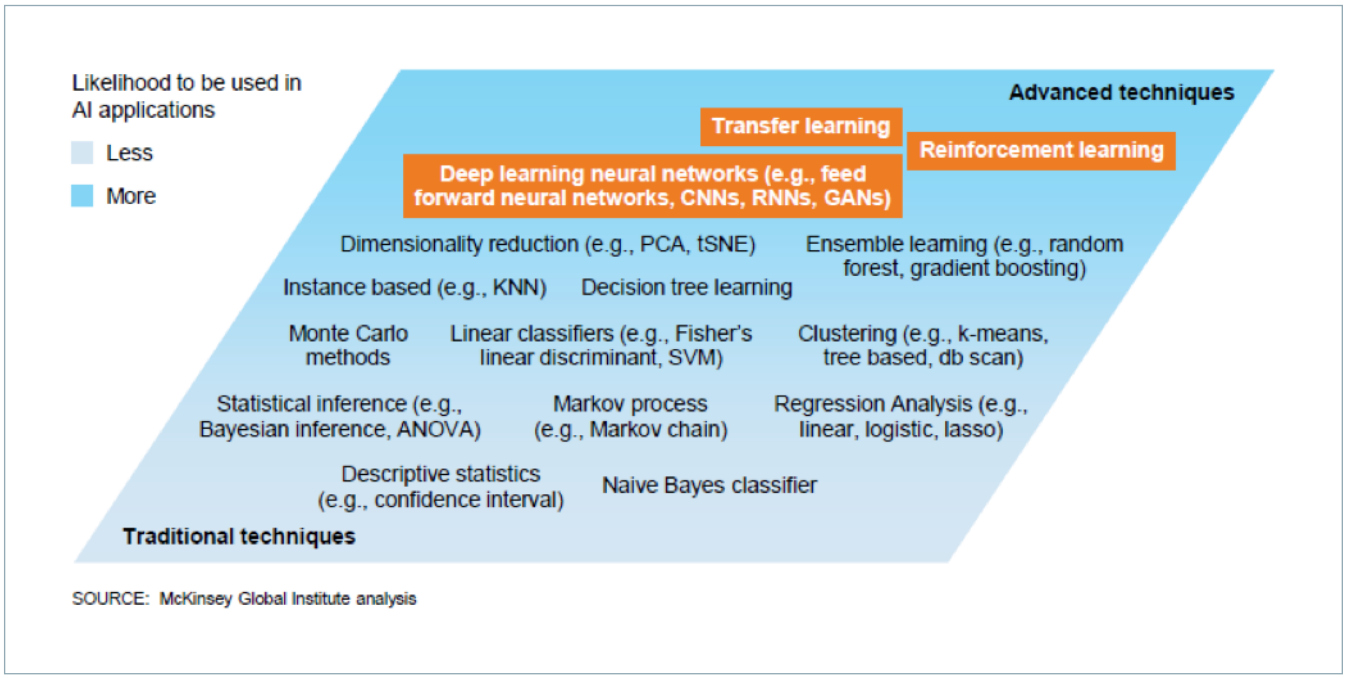
Gartner Research has created a working definition of the concept of "advanced analytics":

> *Advanced Analytics is the autonomous or semi-autonomous examination of data or content using sophisticated techniques and tools, typically beyond those of traditional business intelligence (BI), to discover deeper insights, make predictions, or generate recommendations. Advanced analytic techniques include those such as data/text mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing, neural networks.*[15]

For the purpose of this study, TEConomy considers the techniques and methodologies included across a broad continuum of technological sophistication to be included in the definition of advanced analytics, but with a primary focus on the use of analytics tools and platforms within **applied, industry-facing contexts** to drive process automation, advanced insights, or other value-added tasks in products and services being advanced by companies.

The McKinsey Global Institute provides a useful summary graphic illustrating the scope of the continuum of analytical techniques noted above, from the "traditional" to the "advanced" (Figure 3).[16] For the purposes of this TEConomy study, we are predominantly interested in potential implementations of analytical techniques towards industry needs rather than any one particular segment of advanced techniques. However, special recognition should be given to any applications being advanced that leverage the upper levels of analytical techniques shown below, as they represent more powerful approaches to applied analytics that are often referred to as "artificial intelligence."

Figure 3: McKinsey Continuum of Analytical Techniques



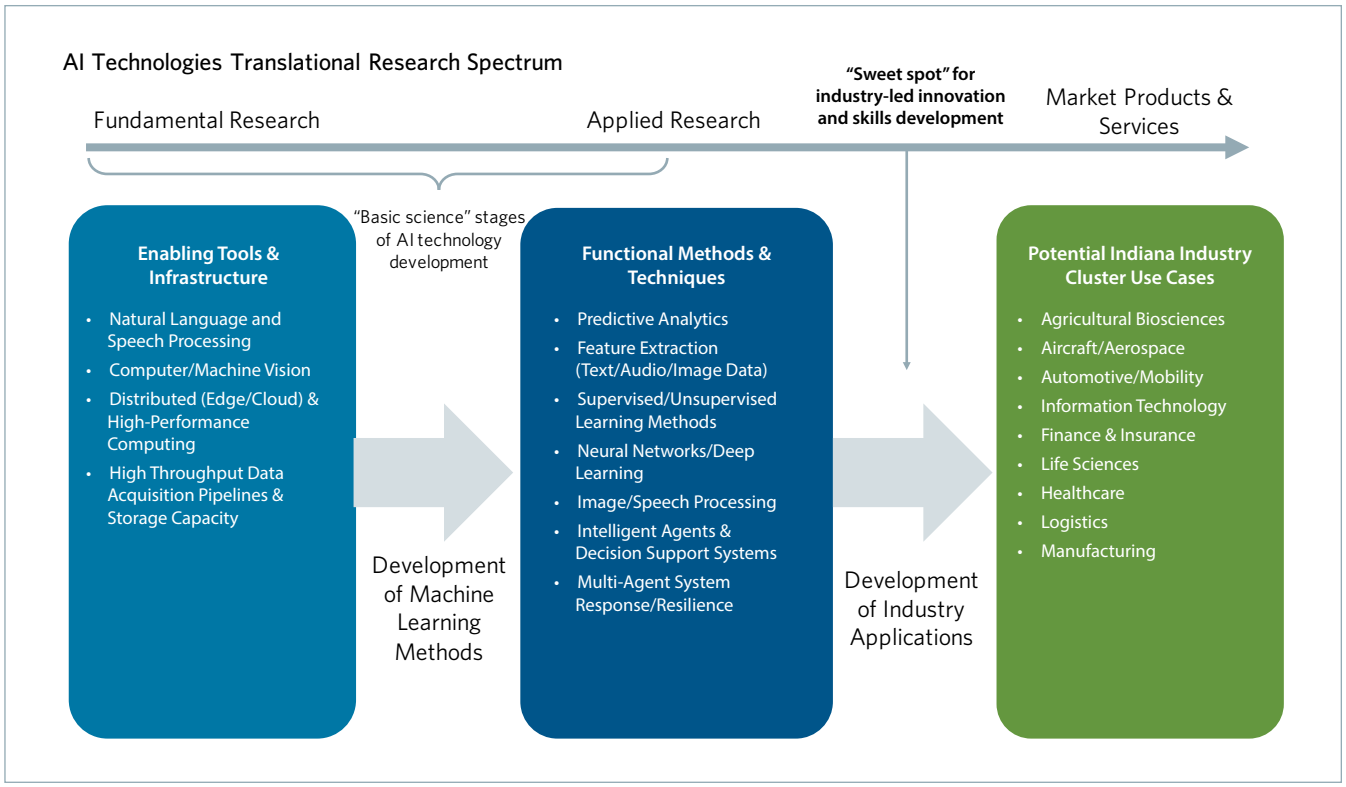
In the case of this study for BioCrossroads, it should also be noted that we are generally more interested in the **application** of advanced analytics more so than in the fundamental research that underpins AI and associated techniques

15 *https://www.gartner.com/en/information-technology/glossary/advanced-analytics.*

16 *Michael Chui, James Manyika, Mehdi Miremadi, Nicolaus Henke, Rita Chung, Pieter Nel and Sankalp Malhotra. "Notes From the AI Frontier: Insights from Hundreds of Use Cases." Discussion Paper. McKinsey Global Institute, April 2018.*

development. However, given the speed in which the science is advancing, applications of advanced analytics are much closer to open research questions than traditionally found in other fields of inquiry. As Figure 4 illustrates, there is a research and development continuum that runs from fundamental computer science, statistics and associated inquiries in the development of new techniques and tools for data analytics through to the application of established techniques to applied analytical and automation needs derived from business problems or creation of goods and services. **This project primarily concentrates on evaluating the current status of the State in advanced data sciences as they relate to opportunities that are of relevance to economic futures.** That said, fundamental research can still be a basis for development of commercializable algorithms and tools and does receive consideration in our analysis.

Figure 4: Structure of the Translational Research Spectrum for AI-Enabled Technologies: Tools, Methods, and Applications



F. About This Study

Technology-based economic development (of which AI and advanced analytics development and deployment are a part) is empowered by ecosystems comprising R&D, talent and capital. Consideration of Indiana's position and promise in advanced analytics and AI requires that the current situation across an AI-enabling ecosystem be understood.

BioCrossroads and its stakeholders certainly have an awareness of the potential importance of advanced data analytics to the future success of the Indiana economy. What is less clear for BioCrossroads, however, is the current status of the state in advanced data sciences and where there are strengths, weaknesses, opportunities and threats related to data analytics that are of relevance to the state's economic future.

BioCrossroads identified that it needed answers to central questions such as:

- Where do the state's higher education institutions stand in terms of R&D and educational core competencies in major aspects of advanced data sciences?
- In what specific areas of current and emerging advanced data analytics R&D does the state have a competitive advantage, and a line-of-sight to significant markets, based on R&D core competencies?
- Which industries in the area are currently applying advanced data analytics and which are on the frontiers in terms of application of highly advanced techniques in machine learning, AI and associated fields?
- Which areas of advanced analytics offer an opportunity for Indiana to "go big" in terms of building upon R&D core competencies AND providing relevance to the ongoing economic performance of key advanced industries?
- What large-scale institutional data-sets exist that may be better leveraged as assets via the use of AI analytics (for example, Indiana Health Information Exchange, Regenstrief Institute, etc.)
- What significant gaps exist in core competencies in advanced data analytics that need to be addressed?
- How well is the state performing in terms of talent generation and attraction relevant to advanced analytics disciplines and business functions?
- What are the entrepreneurial strengths around the development and deployment of new advanced data analytics companies and services, and how can Indiana distinguish itself as the smart place to fund and build start-ups in this field of growing opportunity?

In providing answers to these and other project questions, TEConomy deployed the following methodology (Figure 5):

Figure 5: Basic Steps in Project Performance

The chapters that follow in this report present the findings from quantitative and qualitative evaluation of Indiana's current position in advanced analytics and AI. Chapter II presents an industry perspective, covering findings and conclusions from a series of interviews with major employers (predominantly comprising companies in advanced industries). Chapter III profiles the supporting ecosystem for advanced analytics in Indiana, reviewing:

- Assets and advanced analytics-focused programs deployed, or in development, at Indiana's research universities;
- Occupation and talent supply in associated occupations and fields;
- Entrepreneurial activity in advanced analytics and AI; and
- Organizations, programs and events in support of advanced analytics development in Indiana.

Chapter IV presents strategic recommendations for Indiana resulting from the full program of work.

II. Industry Needs and Opinions Regarding AI and Advanced Analytics in Indiana

A. Input from Major Employer Interviews

In order to directly capture industry stakeholder viewpoints on the current state of AI-enabling capacities and needs, TEConomy conducted in-depth interviews with major employers located in the state, particularly those operating in Indiana's advanced industries. These companies spanned the major industry clusters present in the state, including, for example, biomedical, manufacturing, and financial services sectors. Interviews were conducted as a means of intelligence-gathering around general themes, with no attribution back to individual companies or interviewees.

Companies were asked questions on several key topics that sought to build a holistic picture of their position, capacity, and level of activity in advanced analytics technologies, including:

- Current business activities related to analytics and AI technologies;
- Present and anticipated future business needs leveraging analytics and AI technologies;
- Interaction and collaboration with universities and other research institutions, both in-state and nationally;
- Thoughts on the ability to source and integrate skilled analytics talent into company operations including desired skill sets, attraction, recruitment, and retention of talent, relationships with talent-generating institutions, and reskilling/upskilling of the current labor force;
- Recommendations for growing company and broader Indiana capacity in AI-enabling technologies and talent; and
- Specific use cases for advanced analytics technologies that would have significant short-term value for companies.

Feedback from the industry interviews has been summarized below across general themes.

1. High Demand for Building Analytics Capacity, but Difficulty Bridging the Gap to Applications

The Indiana companies interviewed as a part of this study universally understood the importance of analytics to modern business models, and generally felt a sense of urgency to understand and incorporate these technologies into their business operations. However, there is a gap in the general knowledge and implementation roadmaps required to help them bridge the gap from their current state to a fully realized analytics capacity before value can be realized within the company through downstream applications. Increasingly, the companies are making significant investments in order to bridge these gaps using expert consulting resources and establishing operations with a focus on specialized technical talent in analytics hubs in California, Massachusetts, Toronto and elsewhere. Companies highlighted the complex nature of growing an advanced analytics ecosystem, with intertwined needs across back end IT infrastructure, expertise in developing solutions, and creation of applications and end products that can provide meaningful insights, with many citing the lack of centralized repositories of information and guidance in building out their capabilities. Companies also expressed a desire for solutions to move "at the speed of business" in terms of flexibility and

responsiveness to internal company pressures, something that was typically lacking in off the shelf analytics solutions and available skill sets.

2. Broad and Highly Company-Specific Spectrum of Needs

Companies expressed a wide variety of levels of technological readiness and maturity with respect to their capacity and needs in the space of advanced analytics and AI technologies. While some companies expressed a desire to deploy high end data processing and modeling capabilities to help advance specific segments of their business, many companies are still at nascent stages of beginning to incorporate basic analytical approaches and back-end data gathering and retrieval infrastructure into their operations. This variation in technological readiness spanned different industry sectors and company sizes, bringing an additional variable that strategic initiatives need to account for in meeting companies at their current state rather than an assumed level of readiness.

Additionally, companies expressed a variety of different levels of desire to invest and scale deployment of these technologies in the short term. Some companies expressed a preference for proof of concept solutions that could demonstrate value and then transition to longer term investment in upgrading capabilities, while others identified the need for enterprise level transformations that would require significant upgrades to both infrastructure and staff. As with the level of technological maturity, there was no consistent clustering of desired scales of deployment across industry sectors or company sizes, again highlighting the very customized nature of solutions development within individual companies, even those working within the same market spaces.

Companies often pointed out that the advanced analytics space was very dependent on interlinked systems and technologies, and that a broader rather than narrower field of view towards developing capabilities was critical. In particular, several companies noted the importance of Internet of Things (IoT) technologies in generating data streams which could then be leveraged towards solving business problems and that this technology needed to be developed alongside analytical capacity in order to drive value. This was a particular point of emphasis for companies in production and manufacturing sectors, where large scale deployment of instrumented equipment was viewed as an integral part of long-term success in upgrading analytics capabilities.

3. Getting Connected with Talent

The most frequent area of concern expressed by Indiana companies involved accessing and retaining skilled talent bases in this area. Almost all companies expressed that they were currently not able to meet internal demand for talent to grow capacity, but the reasons varied by company and industry sector. A number of stakeholders indicated that they have multiple open positions related to data analytics and AI-enabling roles but are not finding the right kinds of talent to fill the positions. A rough estimate of the volume of demand currently is at a volume of several hundred additional data sciences professionals across the cohort of major companies in Central Indiana, and the need will only continue to grow as additional late adopter businesses transition to new operations models.

A number of different reasons across companies were identified that contribute to difficulty in meeting talent needs:

- Attracting talent from outside the state is challenging because skilled professions will often not uproot from east and west coast hubs to come to Indiana for a wide variety of reasons that vary by industry and type of data sciences professional segment. This presents a substantial challenge for the large companies that can drive recruitment from name recognition, let alone for smaller firms or newer ventures. Several companies also noted that their fields of business were not viewed as attractive to skilled talent because of a perception of uninteresting or routine work with established business models or regulatory constraints.
- Industry has highlighted difficulties in recruiting both senior talent to lead and develop internal subject matter expert teams as well as a perception that more entry level talent coming out of regional pipelines does not have

the right applied skills and business process knowledge. There is also a perception that in-state students are hard to retain.

- Companies expressed concern that there is the potential for ongoing problems with regional companies poaching talent from one another, rather than building out a larger talent base, with some evidence that this is an issue in the current ecosystem where top level talent is at a premium.
- Although most of the companies already have skilled STEM labor forces of engineers and scientists (people with talent in mathematics and an aptitude for understanding data), companies noted that it is still a heavy lift to train these individuals in-house to be data scientists due to the mental models built off of their education and past experience. It is perceived to be easier to take a data scientist and impart company subject matter expertise and business practices than vice versa.
- There was some hesitancy expressed by companies to invest heavily in bolstering talent pipelines because of a perception that high performing individuals often use their first job at a regional company as a "springboard" to get experience that enables them to land "better" positions at higher profile companies, often out of state. Many companies felt long term strategic and creative thinking around incentives to root cohorts of talent in Indiana are necessary to address this issue.
- Some companies expressed needs for very specialized talent pools that are rare even on national scales, making competition for these roles difficult for Indiana to engage in. One notable example is in biomedical industries, where the "physician data scientist" is in high demand but due to educational and industry culture norms is very rare to find in practice.
- Companies noted that it is difficult to compete with salary offers from coastal hubs that often attract top talent, and that even with lower in-state cost of living the salary levels they are able to support are often not competitive with other offers top talent is receiving.

Because of the pent-up demand for talent, companies noted that they are often having to triage which areas of their business will be first in line to be internal "customers" for developing solutions and applications driven by advanced analytics. Several larger companies noted that multiple areas of each company want to have the capabilities deployed over the short term, but there is not enough internal capacity to meet all demands. This in turn drives pressure from management to ramp-up capacity rapidly, which some companies have done by turning to other parts of the country or offshore international talent pools. Companies noted that outsourcing analytics needs was not an optimal solution and had already begun to show decreasing returns on investment as labor markets in those regions or countries have begun to increase their own demand for skilled talent.

Most companies described a process of developing capability around an internal "centers of excellence" model composed of a larger scale group of talent that the rest of the business can leverage for solutions as opposed to embedding data scientists into individual business units. This was largely due to the inability to scale embedded data sciences groups using recruitment and also a means of attracting skilled talent into a "cluster" of internal data scientists that present the working environment that top talent typically prefers. Companies also described the need to effectively recruit thought leaders into the company to help oversee talent pools and accommodate their desire not to be involved in management roles, which has resulted in ongoing internal discussions about traditional corporate roles and the need for creation of nontraditional job categories.

4. Current Industry Interactions with Regional Universities

The levels and types of interactions with Indiana universities varied by company, with larger companies in major industry clusters tending to have higher levels of ongoing interaction around analytics and AI capabilities with major schools. Relationships tended to be coordinated and maintained at the individual company level, and historically connected on the university side typically through key entrepreneurially-focused faculty rather than at higher administrative levels. Several companies noted that this has begun to shift over the past several years, as universities such as Purdue have begun to coordinate programmatic efforts to engage companies across administrative units rather than within.

Companies noted that they generally did not have high levels of awareness of faculty that are active in AI-related applications, particularly in applied disciplines that would be of most interest to furthering industry-funded innovation. Many noted that there is currently no well-established mechanism for engaging with the wider body of in-state universities around ideation, joint work, student engagement, and other topics of interest. Many industry stakeholders pointed out that they did not feel a high degree of connectivity to in-state university programs in these areas outside of periodic recruitment of graduates, and that this was likely a contributing factor in difficulties in retaining skilled talent in-state.

Industry representatives who have expressed interest in working with universities note that historically, discussions about increasing engagement have involved the need for significant investment in named university centers at individual universities, which is not aligned with the short-term needs of companies. Companies instead tended to prefer multi-university engagements as well as a broader group of involved industry members that could help sponsor ongoing collaborations. However, companies noted that a one size fits all model of collaboration was also not appealing and that engagements needed to have a consultative nature that could help in addressing specific company problems that could demonstrate value or rapidly scale up talent.

More detailed examples of specific Indiana research university programs engaged with industry can be found in Chapter III Section B.

B. Potential Demonstration Cases Identified by Employers in Indiana

Research universities represent a potential partnering resource for solving both near- and longer-term corporate challenges and needs in advanced data analytics. With potential partnerships in mind and to inform strategic recommendations, discussions with many of Indiana's large and leading corporations included requesting from them specific problems where they are looking to advanced data analytics capabilities for solutions, and for which a research university may represent a valuable partner. These were positioned with each company or organization as potential "use cases" for beginning a dialogue with university leadership and individual researchers that gets more specific and begins to make more tangible those challenges corporations face today and expect to tackle in the coming years. This section highlights these potential demonstration cases/use cases offered by industry.

Recognizing the nature of these areas of focus as highly sensitive and strategic, the use cases presented here are not attributed to specific firms. Rather they are offered to illuminate the types of challenges companies are seeking to solve today, and to inform potential university or private solutions collaborations that fit. Not every company could or would provide use cases for the study (though a majority did), so these represent a subset of the full interview group.

In broad terms, the companies interviewed represent three major industries:

- Advanced Manufacturing,
- Health and life sciences, and
- Insurance.

The use cases offered take on multiple forms—some are industry-, product- or technology-specific; others are focused on enhancing core business and operational functions; and still others from manufacturers are focused on digital transformation of production and maintenance processes. Each use case would draw from one or more specific advanced analytics capabilities and/or technologies as set out in Table 3.

Table 3: Industry-provided Use Cases for Potential Partnering Opportunities with Research Universities and/or Private Industry Solutions Providers

Applications Area	Specific Use Cases Identified	Advanced Analytics-related Competencies/ Tech Required/ Leveraged
Core Business Functions	• **Forecasting Sales:** using advanced analytics to improve forecast accuracy of sales to inform and adjust manufacturing, and to configure supply chain. • **Decision Support for Managing an R&D Portfolio/Strategy:** for R&D portfolio management, considering the balance of risk, investments and rate of new product introduction – simulation of forward time-based scenarios for decision support. • **Enhancing Customer Service:** measurement, analysis, and prediction of customer experience. What is a customer's actual journey as they interact with the company, what are the drivers of dissatisfaction and delight? Given a specific customer, what experience/journey is likely to achieve an outcome (purchase, retention, promotion). • **Gauging Marketing Effectiveness:** retrospective and predictive analysis of marketing activities/spend/campaigns. What's the value to the firm (sales, brand, etc.); how should the firm tune marketing better for individual interactions (end customers, producers, etc.)?	• Predictive analytics • Modeling and simulation

Enhancing IT/ Data Analytics Infrastructure	• **Upgrading, Scaling IT Infrastructure:** scaling up, upgrading IT Infrastructure including hardware, installing new apps.
	• **Back-end IT Development Areas:**
	  • enhancing data quality, workflows, operations management – create partnerships giving students access to data to build algorithms.
	  • assistance with product apps in areas such as oncology clinical decision support workflow.
		• Data warehousing/ storage infrastructure, architecture

Industry/ Product/ Technology Specific Applications	• **Using NLP Tech to Translate Voice to Text:** utilizing Natural Language Processing technology to capture physician notes; align with an industry partner to identify patient candidates for clinical trials.
	• **Leveraging Genomic Data to Inform Clinical Trials:** partner with a third party to use genomic information from a biorepository to compare against clinical trials information.
	• **Analyzing Social Determinants of Patient Health:** a project to leverage structured data regarding social determinants of care (e.g., obesity, smoking cessation).
	• **Studying Opioid Use:** leveraging health systems data and data from the Indiana Health Information Exchange to understand patterns of opioid use.
	• **Leveraging AI and Machine Learning for biomedical/life sciences applications, including:**
	  • for drug discovery and development,
	  • for digital health, and
	  • algorithms for personalized medicine.
	• **Balancing Risk for Third Party Insurance Sales:** many insurance products for individual sold through institutions such as companies, schools, hospitals etc. How relevant is the institution's reputational/ risk scores to assist and impact individual underwriting?
	• **Predicting Insurance Claims on Initial Indication:** in insurance claims, can prediction of the severity or complexity at first notice of loss become more accurate?
		• Natural Language Processing
		• Data fusion and Big Data analysis
		• Predictive analytics
		• AI and Machine Learning

C. Key Conclusions from the Major Employer Interviews

Taken collectively, the interviews with leading Indiana industry stakeholders paint a picture of a state in transition towards analytics-driven enterprise business models with a critical need for skilled talent to sustain momentum in integrating and leveraging new technologies. For a variety of reasons, connections with talent pipelines in the state are not being grown effectively which has put significant pressure on companies to find short term solutions.

Several key conclusions drawn from leading industry stakeholder thoughts on analytics and AI technology needs, talent demand, and collaborative solutions include:

- **With a few exceptions, the Indiana companies interviewed are still at relatively emerging stages of incorporating advanced analytics technologies into their operations.** Corporate leaders are aware of the urgency of the problem and taking active steps to accelerate progress in scaling up capacity but are only just beginning their transition processes and not yet ready to leverage high end technologies in advanced machine learning and true AI.

- **The most pressing need for companies in the short term is finding ways to scale new capabilities quickly and provide short term value to consumers using applied analytics solutions.** The level of maturity of most corporate stakeholders in this space is such that leading-edge solutions for advanced AI topics are likely not aligned with current business needs – addressing the development of a cutting edge AI ecosystem with associated thought leaders is a downstream problem once more short term needs in analytics integration and talent are addressed.

- **Leaders in investing in and integrating technologies in this space tend be in the manufacturing and production industries** since they have embraced the value-add of data sciences in addressing customer cost drivers. Many of these stakeholders have shifted towards digital operations and the resulting large scale data flows from instrumented equipment require upgraded capabilities to provide the long term quality assurance and monitoring capabilities customers are demanding.

- **Some industry segments that are key to Indiana are lagging behind in this space** due to the nature of these industries and their norms and regulatory limitations rather than capacity issues.

    - A key example is the healthcare delivery space, where industry leaders note that enterprise level change has never historically tended to move quickly in adoption of new technologies or business processes both due to medical practitioner culture as well as regulatory compliance and adverse event risk concerns. These types of limitations are difficult things to navigate around quickly and are influenced by national trends, so these sectors may take longer to mature and lag behind other early movers. The tech culture of "run fast, break things" does not responsibly translate to sectors where life and death decisions are impacted.
    - Another example is the IP security concerns and risk averse business model of manufacturers that run counter to new data sciences paradigms of open collaboration and data sharing, so it may take time to build trust in solutions with those that have not yet made the strategic decision to invest in these capabilities internally.

- **Many industry leaders felt that a significant investment in a signature initiative involving a collaborative group of leading companies and institutions was likely required to demonstrate meaningful progress in advancing the state's current ecosystem.** Companies had mixed thoughts on what form such an initiative should take and what type of location model it should utilize but agreed that university capacity was best suited to helping provide expertise for more advanced topics of interest as well as advancing talent pipeline build out. Companies overwhelmingly agreed that collaborative efforts should be structured to be consultative such that individual companies can introduce specific needs and challenges for customized projects at varied levels of engagement.

- Industry leaders also have **a need for solutions to help up-skill and re-skill their existing workforces in new data sciences methods and tools,** and note that successful talent initiatives cannot rely solely on new graduates to drive change and incorporate business knowledge into new applications.

- Industry leaders noted a lack of major anchoring commercial solutions providers or clusters of significant vendor satellite offices in-place or being built in Indiana that would fill a need for third-party tools, integration, and expertise in best practices. Examples of these areas include cloud computing providers, networking and IT systems integrators, data analytics tools and software vendors, and other developers of hardware and software that supply industry with capabilities to empower skilled analytics workforces. Many companies noted that they are currently dependent on third-party point solutions to bridge the capabilities gap as they scale up their own internal teams, and they are primarily acquiring these products from outside the state. Companies also indicated a lack of awareness of and connection with the small cohort of emerging Indiana-based analytics solutions and integration providers identified by TEConomy, and expressed a desire to learn more about in-state options for procuring services. This ecosystem stands in contrast to robust analytics environments like the Research Triangle region of North Carolina, where companies have large data sciences companies to partner with locally (such as SAS and IBM) and which also serve as attractors for talent that can migrate between companies over time.

- The current situation and needs of industry stakeholders will require a multifaceted strategy that incorporates elements of university-industry collaborative partnerships, industry-to-industry networking and knowledge sharing, talent pipeline capacity building initiatives, an effective strategy for funding and attracting start-ups and building an entrepreneurial supply chain, and development of centers of excellence around which to focus attention and build critical mass. Building capability in advanced analytics requires structuring solutions differently than traditional industry-academic partnerships due to the rapidly evolving and highly applied nature of data sciences.

III. Indiana's Advanced Analytics and AI Ecosystem – Current Status

A. University Research Capabilities and Core Competencies

Data analytics is a field of inherent cross-cutting importance for most fields of scientific and technological inquiry. The tools and techniques of data analytics power advancements across a host of other disciplines, enabling insights to be derived from data that lead to innovations and technological advancements. Universities are a particularly important component of the data analytics ecosystem because they are able to pursue fundamental research inquiry that underpins the development of new insights, innovations and techniques; and comprise an intrinsically multi-disciplinary community able to adapt and test new techniques across a broad range of disciplines and applications. Universities are also the training ground for a workforce with skills honed across both fundamental data sciences concepts and the application of data sciences to particular fields and disciplines.

Developing an in-depth understanding of Indiana's AI and advanced analytics ecosystem thus mandates that an understanding of the core competencies and capabilities of Indiana's major academic institutions in this important cross-cutting field be developed. To that end TEConomy undertook both quantitative and qualitative research to develop a profile of data sciences, AI and advanced analytics capacity at major higher education institutions in the state.

1. At the Core – Analytics within Computer Science

The analysis of digital data, and the development of techniques for doing so, is a field of inquiry with deep roots in computer science as well as in mathematics and statistics. To gain an overview of Indiana's standing in AI and associated disciplines within computer science, TEConomy accessed the dataset compiled by CSRankings – a web-based data resource developed by Dr. Emery Berger at the University of Massachusetts Amherst.[17]

The CSRankings data resource is best viewed as identifying a core set of computer science (CS) and associated faculty who are engaged in various subject matter areas within CS. It is a uniquely effective tool for identifying people at the core of the disciplines. It is less effective, however, at identifying people at minor institutions or in peripheral (non-CS) areas that may have AI engagement. CSRankings was designed to primarily serve prospective graduate students who are seeking a postgraduate degree in Computer Science – helping them compare institutions and identify faculty working in areas of interest to the student. As such, the CSRankings data are well structured for enabling a comparative overview to be performed of institutions and for developing comparative state profiles. **CSRankings is entirely metrics-based:** it weighs departments by their presence at the most prestigious publication venues – an approach that is intended to be both incentive-aligned as faculty already aim to publish at top venues and "difficult to game", since publishing in such conferences is difficult.

17 CSRankings developer, Dr. Emery Berger (Professor of Computer Science at UMass Amherst), kindly provided permission for the use of the data for this analysis for Indiana.

CSRankings notes that:

- Almost all categories used in the rankings are based on research-focused Association for Computing Machinery (ACM) Special Interest Groups (SIGs). Areas not represented by ACM SIGs are intended to span most established research-centric areas of computer science. For any research-focused area to be included, at least 50 R1 institutions must have publications in the top conferences in that area in the last 10 years. This threshold is to ensure that there is enough research activity in an area to enable a meaningful ranking.

- The conferences listed were developed in consultation with faculty across a range of institutions, including via community surveys. Only the very top conferences in each area are listed. All conferences listed must be roughly equivalent in terms of number of submissions, selectivity and impact to avoid creating incentives to target less selective conferences.

Indiana Institution Positions in CSRankings

The areas of computer science are segmented as follows in the CSRankings data (Table 4):

Table 4: Areas Included in CSRankings

Meta Grouping	Areas Included
Artificial Intelligence	Artificial intelligence, Computer vision, Machine learning & data mining, Natural language processing, The Web & information retrieval
Systems	Computer architecture, Computer networks, Computer security, Databases, Design automation, Embedded & real-time systems, High-performance computing, Mobile computing, Measurement & performance analysis, Operating systems, Programming languages, Software engineering
Theory	Algorithms & complexity, Cryptography, Logic & verification
Interdisciplinary Areas	Computational biology & bioinformatics, Computer graphics, Economics & computation, Human-computer interaction, Robotics, Visualization

Including all areas of computer science (all areas on Table 4), provides the following individual institutional rankings (Table 5), with Purdue University evident in the top 20.

Table 5: Top 20 Institutions as Ranked by CSRankings Across All Computer Science (2010-2019), Plus Indiana Institutions

Rank	Institution	State	Count [18]	Faculty
1	Carnegie Mellon University	Pennsylvania	16.7	159
2	Massachusetts Institute of Technology	Massachusetts	11.8	86
3	University of Illinois at Urbana-Champaign	Illinois	11.0	96
4	Stanford University	California	10.5	64
5	University of California - Berkeley	California	9.6	84
6	Cornell University	New York	8.9	75
7	University of Washington	Washington	8.6	62
8	University of Michigan	Michigan	8.4	74
9	University of California – San Diego	California	7.3	60
10	University of Maryland – College Park	Maryland	7.2	67
11	Georgia Institute of Technology	Georgia	7.0	87
12	Columbia University	New York	6.3	52
12	University of Wisconsin - Madison	Wisconsin	6.3	52
14	Northeastern University	Massachusetts	5.8	63
15	University of Pennsylvania	Pennsylvania	5.4	51
16	University of California – Los Angeles	California	5.3	36
17	**Purdue University**	**Indiana**	**4.8**	**54**
18	University of Texas at Austin	Texas	5.2	44
19	University of Southern California	California	4.7	40
20	New York University	New York	4.5	51
20	University of Massachusetts Amherst	Massachusetts	4.5	45
53	**Indiana University**	**Indiana**	**2.3**	**47[19]**
60	**University of Notre Dame**	**Indiana**	**2.0**	**23**

Taken together, the three Indiana based research universities (IU, Purdue and Notre Dame) have 124 CS faculty.[20] If only faculty engaged in the five areas included under the Artificial Intelligence (AI) meta grouping are included, the rank ordering changes (Table 6)

18 Count = geometric mean count of papers published across all areas.

19 Includes eight faculty at IUPUI.

20 For the purposes of this report, faculty at IUPUI are included in IU numbers. Where faculty at IUPUI are associated with degrees awarded by Purdue University, they are allocated to Purdue.

Table 6: Top 20 Institutions as Ranked by CSRankings in Artificial Intelligence and Associated Fields (2010-2019), Plus Indiana Institutions

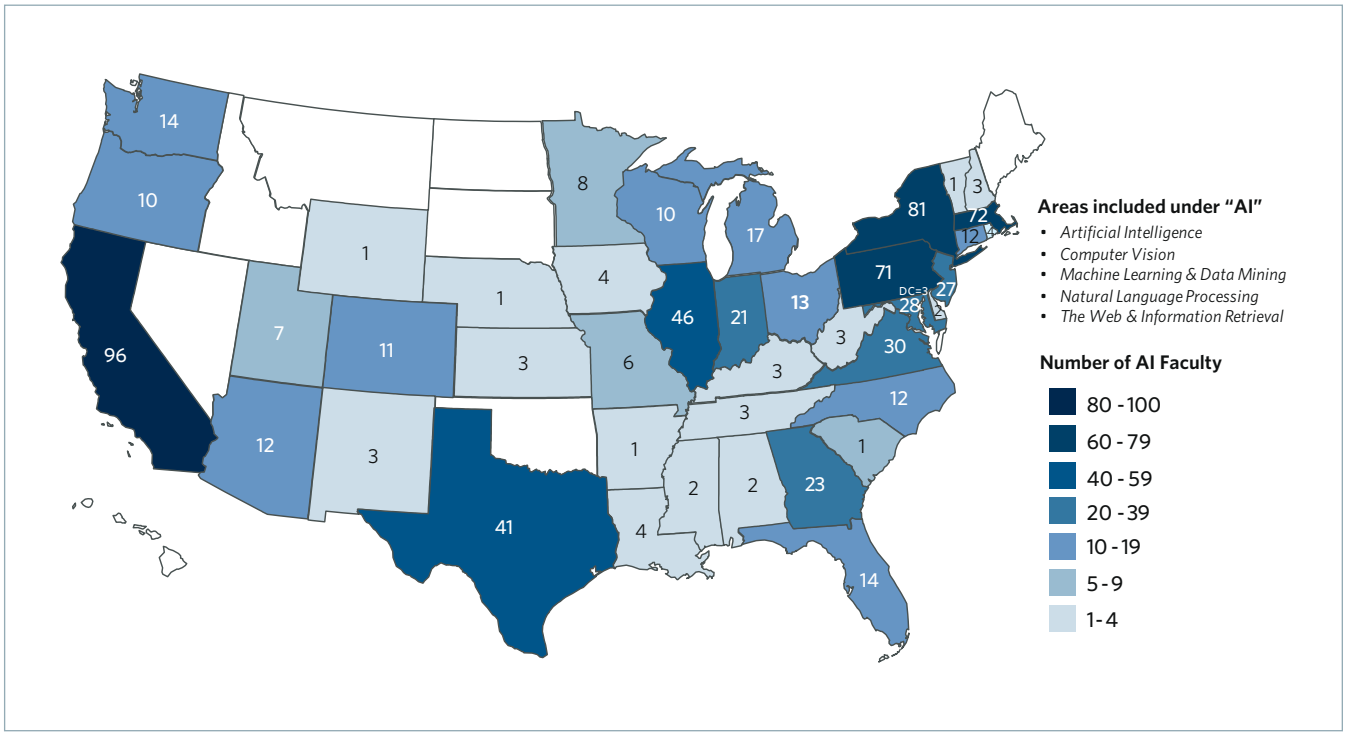
Rank	Institution	State	Count[21]	Faculty
1	Carnegie Mellon University	Pennsylvania	60.3	85
2	Cornell University	New York	34.2	44
3	Stanford University	California	30.6	40
4	University of Illinois at Urbana-Champaign	Illinois	24.2	45
5	Massachusetts Institute of Technology	Massachusetts	20.0	58
6	University of Maryland – College Park	Maryland	18.7	31
7	University of California – Los Angeles	California	17.1	20
8	University of Massachusetts Amherst	Massachusetts	17.0	29
9	Rutgers University	New Jersey	16.9	27
10	Georgia Institute of Technology	Georgia	16.0	41
11	Columbia University	New York	15.5	27
11	University of Michigan	Michigan	15.5	33
13	University of California - Berkeley	California	15.0	44
14	University of Pennsylvania	Pennsylvania	14.9	28
15	University of Texas at Austin	Texas	13.8	22
16	University of Washington	Washington	13.3	32
17	Northeastern University	Massachusetts	12.9	36
18	University of California – San Diego	California	12.4	25
19	University of Southern California	California	12.3	23
20	University of Illinois at Chicago	Illinois	11.8	23
23	**Purdue University**	**Indiana**	**9.3**	**31**
59	**University of Notre Dame**	**Indiana**	**4.3**	**10**
73	**Indiana University**	**Indiana**	**2.7**	**14[22]**

Taken together, the three Indiana based research universities (Purdue, IU, Notre Dame) have 55 CS faculty working in AI and associated fields over the time period 2010-2019. Because faculty may have relocated during the 2010-2019 time period, some individuals may show up counted at multiple universities.

Figure 6 takes the faculty count, just for 2019 (726 faculty nationwide), and maps it as the sum of faculty in AI fields in computer science by state.

21 Count = geometric mean count of papers published across all areas.
22 Includes five faculty at IUPUI.

Figure 6: Artificial Intelligence (AI) Meta-Grouping Faculty in Computer Science by State (2019)



The Top 10 states, in rank order, by AI engaged faculty in CS for 2019 are shown in Table 7.

Table 7: Top 12 States for Number of Faculty in Computer Science Focused on AI Areas

Rank	State	2019 AI faculty in CS	Percent of National Total (n=726)
1	California	96	13.2%
2	New York	81	11.2%
3	Massachusetts	72	9.9%
4	Pennsylvania	71	9.8%
5	Illinois	46	6.3%
6	Texas	41	5.6%
7	Virginia	30	4.1%
8	Maryland	28	3.9%
9	New Jersey	27	3.7%
10	Georgia	23	3.2%
11	**Indiana**	**21**	**2.9%**
12	Michigan	17	2.3%

Indiana ranks 11th among the 50 states (and the District of Columbia) in number of CS faculty focused in AI and associated fields (n=21). It is second among Midwestern states, behind Illinois. Notably, while California is ranked first, it still only contains 13.2 percent of U.S. CS faculty engaged in the space, and is quite closely followed by New York, Massachusetts and Pennsylvania.

In reviewing the CSRankings data, it is evident that, for the size of the state[23], Indiana performs well, ranked 11th, in the AI space within computer science. Each of Indiana's research universities is engaged in the space, with Purdue being a top 25 institution. Combined, Indiana had 21 faculty in CS in AI for 2019, with Purdue having 10, Notre Dame 7, and Indiana University 4. These show that if inter-institutional collaboration and networking can occur in Indiana in AI fields, the state looks comparatively well placed, and certainly competitive for its size.

2. A Broader View – Artificial Intelligence Published Content with Indiana Authors

The analysis of CSRankings data is useful, providing a comparative placement for Indiana just within the realm of AI in the core discipline of Computer Science. **However, as a cross-cutting academic field, advanced analytics and AI research is undertaken in many other academic disciplines outside of Computer Science.** The development and use of AI, in theory and application, occurs not only in computer science, but also in fields such as statistics, engineering, physical and life sciences disciplines, social sciences and business and economics. A broader view of Indiana's engagement in academic research in, or using, AI and advanced analytics can be developed by evaluating research publication output in these fields.

TEConomy used keyword searches incorporating key AI, machine learning, and data analytics terms and concepts, which identified a set of **2,933 research publications with Indiana authors identified from 2010-through early 2019**. These data show:

- Significant growth in AI-related publications volume in the state from 2010-2019 of 122 percent with particularly substantial involvement by Purdue University and Indiana University.
- Activity heavily concentrated within computer science as expected, but a significant volume also occurring in biomedical research fields.
- Activity is not only confined to universities, with non-profit institute and company activity evident in the data also. However, over 95 percent of the publications have university based authors.

The observation that significant AI/Advanced Analytics publishing will occur in disciplines and fields outside of the core of computer science is certainly evident in the Indiana publishing data. Table 8 illustrates this, showing multiple disciplines engaged.

23 Indiana has the 17th largest population among U.S. states.

Table 8: Research Categories for Indiana AI/Advanced Analytics Published Content

Research Category	Number of Publications Included in Research Category
Electrical/Electronic Engineering	578
Computer Science – AI	425
Computer Science – Information Systems	352
Computer Science - Theory	285
Computer Science – Interdisciplinary Applications	242
Automation & Control Systems	112
Imaging Science & Photographic Technology	109
Computer Science – Hardware Architecture	102
Biochemical Research Methods	101
Computational Biology	101
Computer Science – Software Engineering	97
Environmental Sciences	95
Civil Engineering	94
Remote Sensing	94
Statistics	90
Medical Informatics	85

Latent Topic Model Analysis of Indiana AI Research Publications Themes

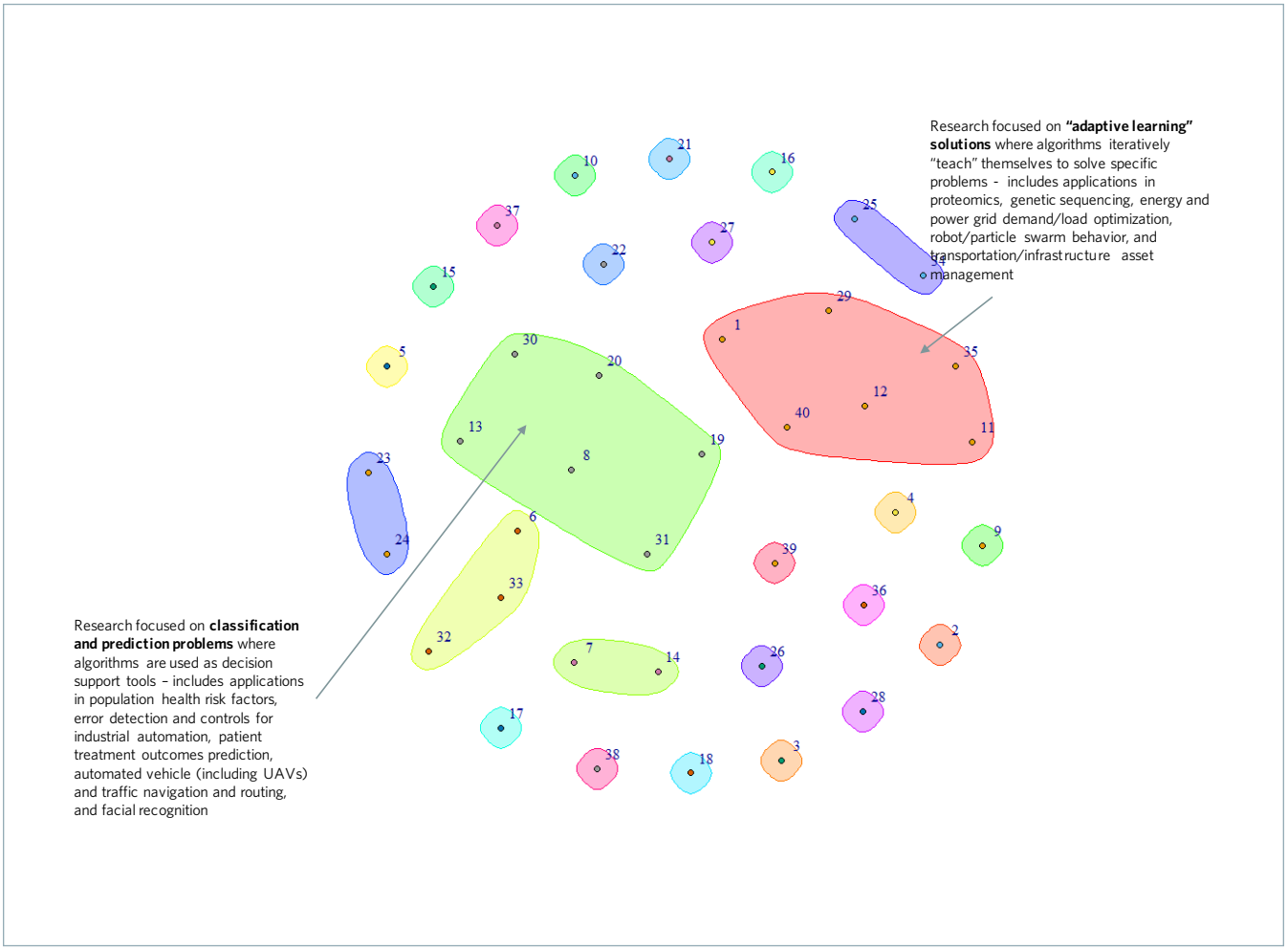
Table 8's classification of research categories is derived from the primary Web of Science classification given to the paper. While that provides a degree of insight, a much more refined picture of AI and advanced analytics publishing in Indiana can be derived by using machine learning algorithms to identify topics or themes that are "latent" within the underlying vocabulary of the text data using a combination of natural language processing (NLP), itself an AI tool, and unsupervised clustering methods. TEConomy thus used descriptive text content from Indiana research publications to form a dataset of unstructured text processed by the analysis.

The latent topic analysis identified 40 topics from the set of 2,933 AI/advanced analytics-related research publications in Indiana that can be grouped into broad research themes.

Findings

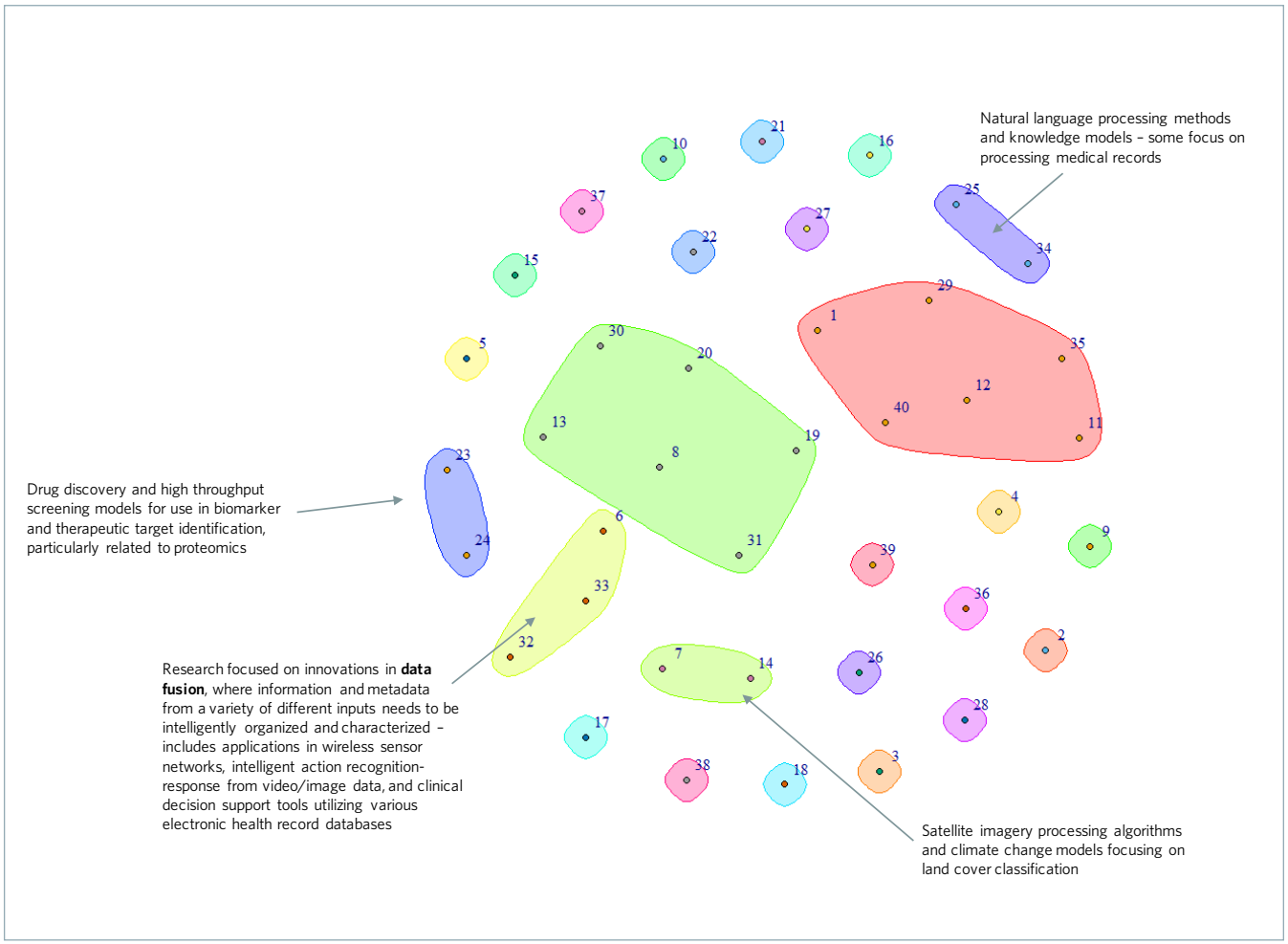
Two research themes displaying critical mass are focused on underlying common techniques/methods used across a variety of specific applications spaces (Figure 7). This demonstrates the cross-disciplinary nature of AI-related innovation and need to specialize in applications areas to differentiate competitive position.

Figure 7: Two Critical Mass Research Themes in Indiana AI/Advanced Analytics Research



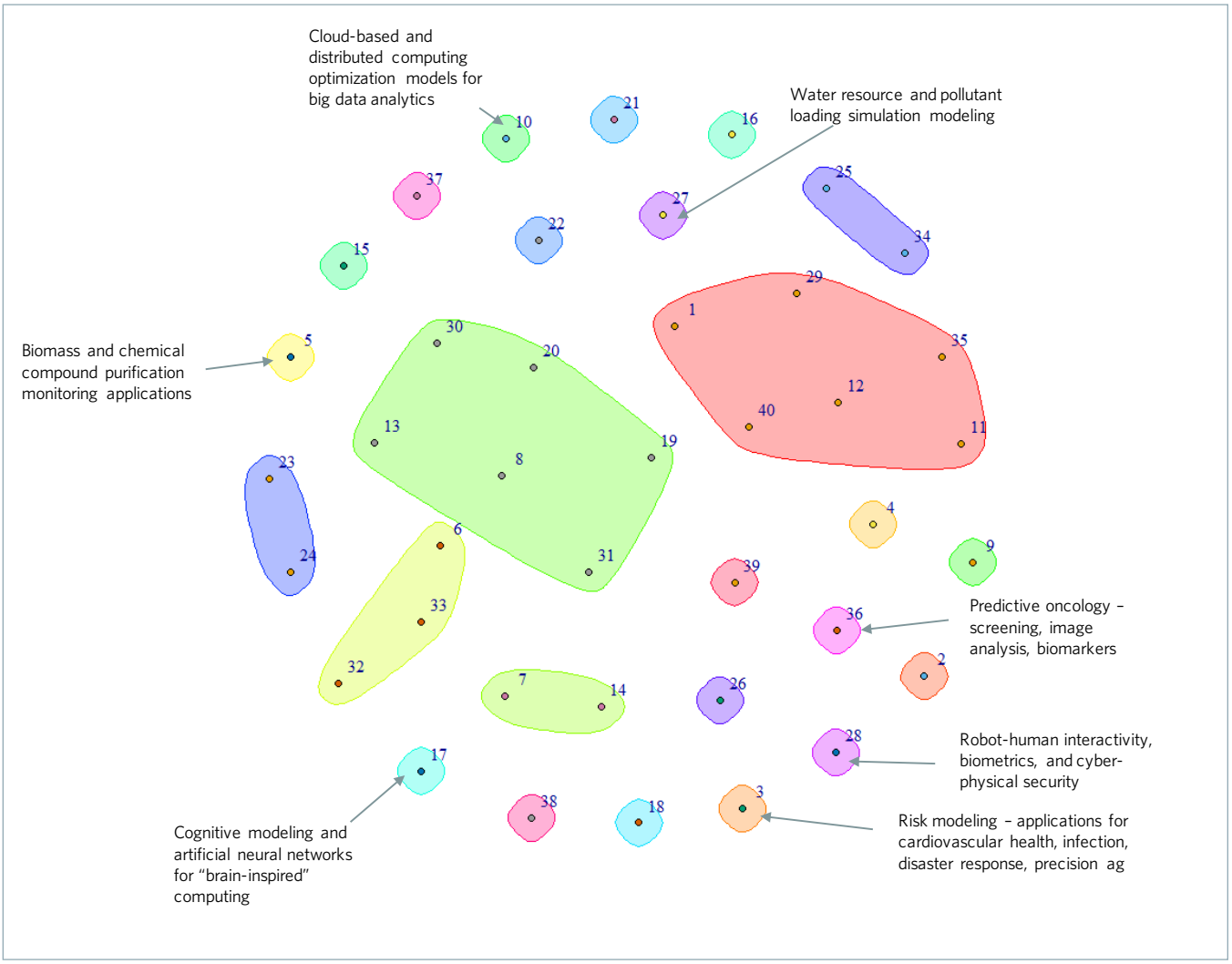
Several other research themes show some levels of distinct activity within publications (Figure 8), indicating potential areas of emerging Indiana specialization for further exploration.

Figure 8: "Distinct Activity" Research Themes in Indiana AI/Advanced Analytics Research



Some other highly focused areas of research that can potentially be leveraged towards novel applications include those highlighted on Figure 9:

Figure 9: Smaller Focused Research Themes in Indiana AI/Advanced Analytics Research



Conclusions from the Publishing Analysis

It is evident that Indiana has recently performed a significant body of research involving the underlying tools of AI-enabled applications, but a few distinct areas of critical mass are evident in differentiated, specialized applications. Research work in the largest activity areas of "adaptive learning" and "classification/prediction techniques related to AI" spans a number of different potential market areas, but several targets for further exploration of alignment to industry are applications focused in biomedical and industrial automation spaces. Another area for further development of strategic alignment is intelligent data fusion, which can have a high value add for industry but needs to be well-positioned for the data gathering needs of specific business models to be useful.

Some areas of focused activity reflect these higher level themes and can potentially help Indiana distinguish itself from competitors in downstream industry-facing applications incorporating AI tools and methods.

B. Indiana's Higher Education Assets and Programs in Advanced Analytics

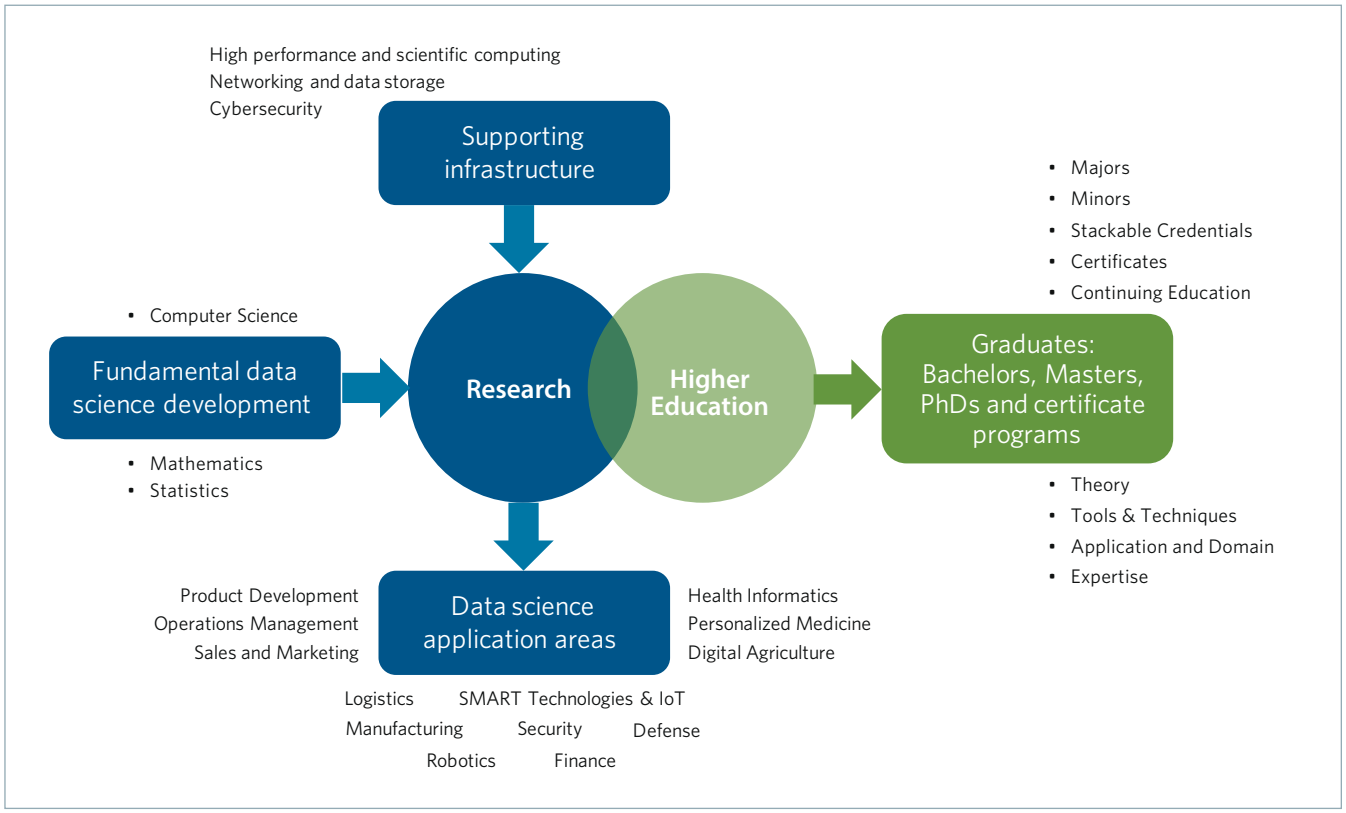
Assessment of publications points to where there are areas of thematic focus within and across Indiana research institutions, however, it provides an incomplete picture in terms of how the institutions are organizing themselves and developing formalized approaches to advance AI and advanced analytics as strategic cross-cutting core institutional core competencies. To investigate this aspect of the advanced analytics ecosystem, TEConomy worked with the universities to perform on campus interviews with university leadership and engaged research groups with an emphasis on organized centers or institutes.

Quantitative methods are central tools for progressing research inquiry in most academic disciplines. Within STEM disciplines, social science disciplines and business studies, quantitative research methods provide critically important lenses for deriving meaningful insight and solutions to research questions. Because quantitative methods are so central to research across a multiplicity of disciplines, they form a cross-cutting competency within universities that are not restrained or contained within any one program, department, school or college. At a major research university, hundreds of faculty with significant expertise in data analytics will be present, and they will be distributed across a broad-range of departments and programs. This is especially true in the case of work in applied data analytics, where researchers are engaged in deploying existing analytical tools and techniques. It is also true, however, even in theoretical and fundamental data sciences, that faculty engaged in leading-edge research can be found within multiple academic homes – in, for example, computer science, mathematics, statistics, economics, informatics and other programs.

Because of the ubiquity of quantitative methods within universities, it is unrealistic to expect to produce a comprehensive overview of all capabilities related to data sciences within major research universities such as Indiana University, Purdue University or the University of Notre Dame. What is possible is to develop an overview of where key clusters of research and education focused around data sciences as a core focus, either theoretical or applied, are evident. On campus interviews and discussions with university leadership, together with review of the academic program websites across the universities, can serve to provide an overview of core data analytics focused initiatives. Findings and conclusions from these evaluations are included in profiles for each of the institutions.

To establish a framework for understanding capabilities at Indiana's research universities, it is helpful to segment activities into typologies. While it is difficult to construct a precise model that fits to the differential organizational structures of individual universities and the continuum of activities that occur, Figure 10 depicts the way TEConomy has chosen to classify different activities at the universities as they relate to observed core competencies and focus areas. It divides activities into two primary domains: "research" and "education", shown as a Venn diagram indicative of the fact that education, especially graduate education, is intertwined with the research mission at research universities. On the research side, three areas are differentiated: "fundamental data sciences development" comprising academic research focused on analytics theory and the development of analytical tools and techniques; "data sciences application areas" comprising the broad variety of applications for data sciences; and "supporting infrastructure" calling out signature capabilities and investments at universities that support advanced data analytics.

Figure 10: Framework for Classifying Data Analytics Activities at Research Universities[24]



1. Indiana University

With over $566 million in research expenditures in 2016, Indiana University ranks in the top quartile of research universities in the nation and has over 200 research centers and institutes.[25] The university's Bloomington campus has approximately 43,000 total enrolled students (2018), operates 550+ academic programs and offers >200 undergraduate majors.[26] Given the scale of IU's research and higher education activity, it is not surprising to find a significant base of assets and programs exist at IU in advanced data sciences and quantitative analytics, spanning the spectrum from theory and fundamental development work to domain oriented application of AI and advanced analytics.

a. IU Fundamental Data sciences Development

In terms of theory and fundamental data sciences development, IU sees significant work taking place in the School of Informatics, Computing and Engineering (SICE). SICE operates a series of data sciences research centers that perform basic theoretical inquiry and work to develop innovative tools, techniques and algorithms for advanced data analytics:

- **Data to Insight Center (D2I)** "focuses on drawing new insights from vast data sets by developing innovative tools and examining the full life cycle of digital data."[27]

24 *Notations on the graphic in terms of data science application areas are not intended to be a complete list of all application areas. Rather a series of example areas of application are listed to show diversity of applications.*

25 *Martin Grueber, Ryan Helwig and Simon Tripp. "Assessing R&D Funding Across Indiana's Major Research Universities." October 2018. TEConomy Partners, LLC. Prepared for BioCrossroads.*

26 *https://www.indiana.edu/about/ranking-statistics.html*

27 *https://datascience.indiana.edu/research/index.html*

- **Center for Algorithms and Machine Learning (CAML)** "gathers research from different SICE departments in order to foster excellence in algorithm development for machine learning."[28]
- **Network Science Institute** "develops foundational network science theories, methods, and analytical tools."[29]

Also of relevance to the advanced analytics, and the associated systems that support advanced analytics and data sciences, are additional SICE centers, including:

- **Digital Sciences Center** which "advances cloud computing and network science, with a focus on developing human-centered interfaces to cyberinfrastructure."[30]
- **Center for Complex Networks and Systems Research**, which "fosters interdisciplinary research in all areas related to complex networks and systems, computational social science, and data sciences.[31]
- **Cyberinfrastructure for Network Science Center**, which "advances datasets, tools, and services for the study of biomedical, social and behavioral science, physics, and other networks."[32]

IU also highlights several labs that are specifically focused in advanced analytics and AI, including:

- **Comparative Classification and Documentation Lab**, which works in theoretical and methodological approaches to the organization of knowledge.
- **Complex Adaptive Systems and Computational Intelligence Lab**, which examines the informational properties of artificial systems that enable adaptation or evolution.
- **Computer Vision Lab**, developing statistical and machine learning techniques for the automated analysis, understanding and organization of visual information.
- **Vehicle Autonomy and Intelligence Lab,** doing work in autonomy and intelligence in robotics systems (for air, terrestrial and aquatic applications).

b. IU Data sciences Application Areas

IU in Bloomington operates multiple research centers directly or tangentially focused on advanced analytics, AI, intelligent systems and associated fields. Many of the centers listed under "fundamental data sciences" above, also undertake applied work. In addition, centers with a distinctive applied focus include:

- **The Center for Bioinformatics Research** undertaking research in evolution and comparative genomics, precision medicine, predictive functional genomics, and structural bioinformatics.
- **Center for Security and Privacy in Informatics, Computing, and Engineering.** Focused on design, evaluation and implementation of technologies that are engaged in the control of digital information and data.

As an inherently applied school, the Kelley School of Business is active in applied business analytics and operates the "Institute for Business Analytics" (IBA). The IBA was one of the first business analytics programs established in the U.S. and includes:

- An academic program preparing students to solve business problems using analytics.
- Corporate partnerships that shape Kelley's understanding of analytics and help companies tap into the School's talent.
- Cross-disciplinary research by Kelley's faculty.
- Seminars, conferences, a speaker series, and a case competition that brings together professionals, faculty, and students.

28 Ibid
29 Ibid
30 Ibid
31 Ibid
32 Ibid

Research by faculty at the Institute for Business Analytics is primarily contained within four focus areas: Finance, Marketing and Retail, Supply Chain, and Healthcare.

c. IU Higher Education

Industry in Indiana cites challenges in recruiting personnel with educational credentials in advanced analytics and AI, yet when examining the major research universities, including IU, it is evident that quite intensive and varied degree and certificate programs exist on campus and online in developing graduates with these skills. Indeed, the volume of graduates being produced at IU, just in business analytics, for example, is considerable with senior leadership at the business school noting that IU is graduating upwards of 600 students a year with business analytics degrees. The IU School of Informatics, Computing and Engineering (SICE) is similarly deeply engaged in advanced analytics education, providing both BS and MS programs in Data sciences, and a minor in Data sciences at the PhD level. Review of related educational programs at IU shows a diversity of advanced analytics options (Table 9).

Table 9: Degree (Undergraduate and Graduate) and Certificate Programs at IU in Advanced Analytics and Associated Fields.

IU School	Degree/Qualification Programs
School of Informatics, Computing and Engineering	• BS in Data sciences • BS in Intelligent Systems Engineering • BS in Informatics • BS in Statistics (Dept. of Statistics in College of Arts & Sciences) • MS in Data sciences (residential and online programs) • MS in Statistical Science • MS in Applied Statistics (for PhD students) • MS in Bioinformatics • MS in Intelligent Systems Engineering
Kelley School of Business	• BS in Business – with Business Analytics Major (second major only) • MBA Business Analytics Major (full-time or online programs) • MS in Business Analytics (online) • Business Analytics Graduate Certificate

IU is proactive in providing students with opportunities to participate in research in their field of study at both undergraduate and graduate levels. An example of this is the Undergraduate Research Opportunities in Computing (UROC) program, to which students may apply multiple times during their time at IU. UROC provides students with a research experience mentored by a faculty member or PhD candidate; and students meet regularly with their mentor, have research team meetings, take a research course and participate in poster sessions.

National data are collected by the National Center for Education Statistics (NCES) from individual institutions on their graduate totals by individual degree fields. NCES uses a common national coding scheme for degree fields that in many cases does not mirror the exact terminology used by the institutions themselves. While the data are not as specific as the programs and degrees presented above, they provide key insights into the scale of the graduate pipeline by individual institution.

Graduate counts are presented in Table 10 for IU in specific degree fields most relevant or "core" to advanced data analytics. **For IU there were 1,522 degree and non-degree award graduates in 2017 in core fields,** the vast majority of which were at the Bachelor's and Master's levels. IU's computer sciences strengths and concentrations are evident in the leading fields of informatics and computer science. The University also graduates significant numbers in economics and math.

Table 10: IU (All Campuses) Graduate Totals in Degree Fields Most Closely Aligned with Advanced Data Analytics, by Degree Level, 2017

Degree Field	Certificates & Awards	Bachelor's degree	Master's degree	Doctorate	Total
Informatics	2	473	7	16	**498**
Computer Science	3	161	208	11	**383**
Economics, General		231	12	11	**254**
Mathematics, General	5	147	11	13	**176**
Computational Science	19		116		**135**
Statistics, General		12	13		**25**
Mathematical Statistics and Probability			19		**19**
Applied Mathematics, General	2	5			**7**
Mathematics and Computer Science			6		**6**
Information Technology	4		1		**5**
Bioinformatics			5		**5**
Computer Programming/ Programmer, General	4				**4**
Artificial Intelligence			3		**3**
Computer and Information Systems Security/Information Assurance			2		**2**
Total	**39**	**1,029**	**403**	**51**	**1,522**

Source: National Center for Education Statistics, IPEDs postsecondary degree database.

d. IU Supporting Infrastructure

IU is exceptionally well-resourced in terms of scientific computing hardware and infrastructure and in advanced high performance network management and operations. Managed by the Research Technologies division of UITS, IU maintains world-class research computing systems including a Cray XC40 supercomputer "Big Red III" and the Big Red II high performance parallel computing system. In addition, IU has the "Carbonate" large memory computer cluster configured to support high-performance, data-intensive computing (with dedicated compute nodes to support deep learning applications and research) and operates KARST (a high-throughput computing cluster designed to deliver large amounts of processing capacity over long periods of time).

The capabilities of IU in high performance computing and network science provide an internal pool of assets and expertise that is able to support highly compute intensive analysis of extremely large-scale datasets.

The IU Campus in Indianapolis and IUPUI

IU maintains a substantial presence in Indianapolis, with the city home to the IU School of Medicine (the largest medical school in the U.S. in terms of enrollment) as well as other university research and education programs. The IU campus in Indianapolis is at IUPUI which was formed in 1968 as a result of the merger of the Indianapolis operations of Indiana University (with IU Indianapolis founded in 1916) and Purdue University (with Purdue Indianapolis Extension Center established in 1946). Degree programs at IUPUI lead to degrees awarded by either IU or Purdue, depending on the program of study.

Today, the Indianapolis campus offers more than 450 undergraduate, graduate, and professional programs from Indiana University and Purdue University. 29,579 students were enrolled at IUPUI for Fall 2018[33], comprising 21,246 undergraduates and 8,333 graduate or professional students. Given the presence of major IU schools, and an intensive series of graduate programs, the Indianapolis campus sustains a significant volume of research. It is home to more than 2,800 faculty located across 17 degree-granting schools and two colleges. The campus includes approximately 100 research centers, including 11 Signature Centers.

As home to the IU School of Medicine and its associated biomedical and life sciences research ecosystem, the Indianapolis campus sees significant research and education activity in these fields. The IU School of Medicine, for example, accounted for more than $355 million in research grants and awards in fiscal year 2018. As would be expected given this research presence, computer science and data analytics core competencies at IUPUI demonstrate a notable emphasis in applications of analytics to biomedical and health sciences data and challenges. This is not exclusively the case (with notable work taking place in other areas such as human-computer interaction, library and information science, and media arts and science), but it is a key characteristic of data sciences and advanced analytics work in Indianapolis. Primarily, informatics programs are provided by the IU School of Informatics and Computing, which was the first school in the nation focused in informatics. The school notes that it:

> *Integrates computing, social science, and information systems design in unique ways to explore how people use computing and technology to live, work, play, and communicate. We apply our insights to developing innovative IT solutions that transform fields like healthcare, biology, business, law, entertainment, and media.*[34]

The informatics program has a robust presence in Indianapolis and combined with IU's other campuses, the School of Informatics and Computing has approximately 2,000 students. Centers and major research programs focused in informatics, include for example:

- The IUPUI Center for Cancer Population Analytics and Patient-Centered Informatics, and
- BioHealth Informatics Research Center – focused on combining data analytics with applied industry engagement to create tools for improving healthcare.

Also of note within the School of Informatics and Computing is the Data to Action (DATA) Lab which comprises an interdisciplinary team of faculty and student researchers performing work to characterize data "not only as a statistical input or technological byproduct, but as a socio-technical construction with inherent contradictions, problems, and ethical implications."

In addition, the IU campus in Indianapolis is home to the **Regenstrief Institute** which draws faculty investigators from the Indiana University School of Medicine, Richard M. Fairbanks School of Public Health at IUPUI, Purdue University, and others who work within three research centers, the Center for Health Services Research, the Center for Biomedical Informatics, and the IU Center for Aging Research. The Institute was founded in 1969 as a medical research and

33 https://www.iupui.edu/about/rankings-statistics.html
34 https://soic.iupui.edu/about/

development organization and is focused on "improving quality of care, increasing efficiency of healthcare delivery, preventing medical errors, and enhancing patient safety."[35] The Institute's Center for Biomedical Informatics applies "health information technology solutions to generate knowledge about health, disease and treatment, help clinicians make optimal decisions, empower patients, and inform healthcare policy."[36] The Center works on clinical applications, computer-based decision support, data mining, advanced analytics, healthcare information standards, and global health.

In terms of education programs at the Indianapolis campus focused on data analytics, undergraduate and/or graduate programs are offered in:

- Bioinformatics,
- Biomedical informatics,
- Data sciences,
- Health informatics,
- Health information management,
- Human computer interaction,
- Library and information science, and
- Media arts and science.

Multiple IU and Purdue undergraduate and graduate degrees are offered at the IUPUI campus that are relevant to the data sciences, advanced analytics and AI focus of this BioCrossroads study (Table 11).

Table 11: IU and Purdue Degree and Certificate Programs Granted Through Study at IUPUI in Indianapolis in Advanced Analytics and Associated Fields

Degree or Credential	Programs
Bachelor of Science Degrees	Applied Data and Information Science Biomedical Informatics Health Information Management Informatics Full Stack Developer (Informatics and Media Arts and Science double major)
Certificates	Applied Data sciences Applied Information Science Human-Computer Interaction Legal Informatics Medical Coding
Bioinformatics Graduate Programs	Bioinformatics Master of Science Bioinformatics PhD Bioinformatics PhD Minor Biomedical Data Analytics Certificate Omics Technology and Precision Medicine Graduate Certificate Data Sciences

35 *https://www.regenstrief.org/about-us/our-history/*
36 *https://www.regenstrief.org/centers/center-biomedical-informatics/*

Degree or Credential	Programs
Graduate Programs	Applied Data sciences Master of Science Data sciences PhD Applied Data sciences PhD Minor Sports Analytics Master of Science
Health Informatics Graduate Programs	Health Informatics Master of Science Health and Biomedical Informatics Ph.D. Health Informatics PhD Minor Clinical Informatics Graduate Certificate Health Information Management and Exchange Graduate Certificate Health Information Security Graduate Certificate Health Information Systems Architecture Graduate Certificate
Human-Computer Interaction	Human-Computer Interaction Master of Science Human-Computer Interaction PhD Human-Computer Interaction PhD Minor Human-Computer Interaction Graduate Certificate

The federal NCES graduate data highlight varied educational focus areas related to advanced analytics, with no one field dominating. Of the 517 degree and non-degree award graduates in 2017, the majority are in computer science fields in programs such as computer and information sciences and informatics (including medical). The biostatistics and medical informatics programs reflect the applied strengths given the co-location with the IU medical school.

Table 12: IUPUI Based Student Totals in Degree Fields Most Closely Aligned with Advanced Data Analytics, by Degree Level, 2017. Degrees awarded are Indiana University degrees or Purdue University degrees

Degree Field	Certificates & Awards	Bachelor's degree	Master's degree	Doctorate	Total
Computer and Information Sciences and Support Services, Other	5	91			**96**
Operations Management and Supervision	20	45			**65**
Computer and Information Sciences, General			53		**53**
Computer Science		50		3	**53**
Mathematics, General		37	14	1	**52**
Informatics	8	26		7	**41**
Economics, General		29	10	2	**41**
Medical Informatics	6		29		**35**
Computer Engineering, General		23			**23**
Bioinformatics			20		**20**
Computer Engineering Technology/Technician		12			**12**
Geographic Information Science and Cartography	8		2		**10**
Biostatistics				8	**8**
Computer and Information Systems Security/ Information Assurance	7				**7**
Computer Systems Analysis/ Analyst	1				**1**
Total	**55**	**313**	**128**	**21**	**517**

Source: National Center for Education Statistics, IPEDs postsecondary degree database.

2. Purdue University

Research expenditures for 2016 at Purdue totaled $606.3 million, placing the university in the top quartile of research universities in the nation.[37] Purdue University had a Fall 2019 enrollment of 65,718 students across all its campuses, and 44,551 students at its central campus in West Lafayette. The University has 2,480 tenured and tenure track faculty. Over 200 undergraduate majors are offered by Purdue.[38]

As one of the nation's premier land-grant universities, Purdue has a long-standing track record of higher education, research and translational extension activity across agriculture, engineering and scientific disciplines. With both broad and deep programs in STEM disciplines, Purdue naturally sustains robust capabilities and capacity related to advanced data sciences and analytics – from both the perspective of fundamental research and applied R&D.

Purdue explicitly recognizes that the ubiquity of data leads to impacts on all sectors of commerce and society in the U.S. This recognition of data as a cross-cutting theme impacting all has led to the University considering how best to approach integration of data sciences across the University's highly diverse curriculum. In effect, Purdue is explicitly recognizing that digital skills and data analytics represent a foundational skill for student success. **At Purdue, data sciences literacy is seen as being a cross-cutting competency that almost every student should acquire during their higher education.**

Integrating data sciences across a diverse university of the size and scale of Purdue is no easy task and the University has established a purpose designed initiative to assure purposeful implementation is achieved. The **Integrative Data Science Initiative (IDSI)** is at the heart of Purdue's deliberative strategy. The vision of IDSI at Purdue is:

> *To be at the forefront of advancing data science-enabled research and education by tightly coupling theory, discovery, and applications while providing students with an integrated, data science-fluent campus ecosystem.*[39]

Research at the Integrative Data Science Initiative is being structured around both basic and applied inquiry. This is evident in the Initiative noting that it has three macro-level thrusts in research:

- Data sciences Fundamentals,
- Data-Driven Discovery, and
- Data sciences Applications.

37 *Martin Grueber, Ryan Helwig and Simon Tripp. "Assessing R&D Funding Across Indiana's Major Research Universities." October 2018. TEConomy Partners, LLC. Prepared for BioCrossroads.*

38 *https://www.purdue.edu/purdue/academics/index.php*

39 *https://www.purdue.edu/data-science/about/index.php*

a. Purdue Fundamental Data sciences Development

As noted above, the IDSI at Purdue includes "Data Science Fundamentals" among its core focus areas. IDSI summarizes the direction of its work in this theme as follows:

Fundamentals, Methods and Algorithms at Purdue	"Data Science research on fundamental aspects of models, methods, and analyses at the Center for Science of Information (CSoI) focuses on core concepts of information, knowledge, learning, fairness, trust, risk, collusion, privacy, and information-efficient computation. Core research topics include development of suitable models for data, generation and testing of hypothesis; characterization of bounds (limits) on important parameters such as learning, transferability and generalizability, control, computation complexity, and statistical significance; novel methods for construction of learning models, reasoning and explainability, privacy, and fairness; and new methodologies for validation and verification, taking into account tradeoffs between computation, accuracy, and time; and novel applications in adversarial/ game theoretic settings. The Center also addresses issues at the interface of data sciences, systems, security, and privacy, investigating problems in real time learning and control for IoTs and cyberphysical systems, loss of privacy from analytics, techniques for secure analytics, data and computation outsourcing, and valuation."[40]

Some affiliated centers of IDSI engaged in fundamental research in, or relevant to, advanced analytics and AI include:

- **Center for Brain-Inspired Computing (C-BRIC) Enabling Autonomous Intelligence.** This is a well-resourced program supported by $27 million in funding from industry and DARPA.
- **Center for Education and Research in Information Assurance and Security (CERIAS).** Purdue notes that this center stands among the world-leading centers in information security and the protection of critical computing and network infrastructure.
- **Purdue Center for the Science of Information,** working to advance quantitative data representation, communication and processing of information for biological, physical, social and engineered systems.

b. Data sciences Application Areas

Initially, the IDSI is encouraging applied advanced analytics work across three primary applied themes (Table 13):

Table 13: Purdue Integrated Data Science Institute – Primary Applied Themes

Theme	Description[41]
Ethics, Society and Policy	"Much of the work on data science focuses on exciting new technology and other STEM advances, but many of the technological advances will go nowhere, or will fail to reach their full potential, because of the role of policy, law, social norms, consumer behavior and other human factors. Algorithmic ethics are a prime example of this question, with consumers and activists raising the alarm about biases inadvertently reproduced or even strengthened through machine learning. Realizing the full potential of data science requires bringing in these ethical and social issues from the start. In addition, adapting machine learning and other big data analytic techniques for social science will open up exciting new areas of study and will give us fresh insight into human behavior. It will enable us to test new questions and to revisit old, enduring questions with new data. More generally, ethical, legal and social issues (sometimes called ELSI) are currently at the very center of conversations about how to move technological development forward, from drones to automated vehicles, so when Purdue leads on these issues, it will strengthen Purdue's focus on excellence in the STEM fields."

40 https://www.purdue.edu/data-science/index.php
41 https://www.purdue.edu/data-science/index.php

Theme	Description[41]
Healthcare	"Healthcare organizations like the Regenstrief Center for Healthcare Engineering (RCHE) develop and apply data science methodologies to make use of the wealth of different sources of health information to create evidence-based approaches to personalized medicine, improve care pathways in delivery, and empower individuals to effectively participate in their health and wellness."
Defense	"Advances in Data Science hold promise to dramatically improve resource efficiency and effectiveness in accomplishing defense and security missions. To unlock the promise, several challenges must be addressed. For example, how to optimize data gathering and structuring to determine what is the most valuable data to collect and/or ignore? Or, how to maximize human-machine symbiosis via combinations of novel data representations, algorithms, visualizations, and computing architecture that enable never-before seen functionality on small, edge devices and platforms? And, in all cases, how to proceed in a way compliant with moral and ethical standards and policies that ensure this compliance."

Because of the importance of data sciences and advanced analytics to commercial sector advancement, **the IDSI has been specifically structured to encourage and support collaborative university-industry R&D, educational partnerships and consulting services.** *Convergence in Discovery Park*, opened in the fall of 2019, offers space to accommodate IDSI Corporate Partners, as well as space for start-ups, research teams of Purdue students and faculty, and shared co-creation areas. In addition, the existing "Foundry" incubator provides services to help with entrepreneurship. IDSI will be helping industry meet needs for talent and expertise through:

- Interaction opportunities with undergraduate and graduate students through data sciences internships and co-ops, fellowships, case competitions, campus job fairs, minority programs, and other initiatives.
- Facilitated collaborations with data sciences researchers at the University through membership in the key data sciences related centers.
- Joint university-industry basic research on common interests defined by the partners, with preferential access to any resulting IP.
- Ability to work with university experts on private applied research on topics defined by the partner, with IP only available to the partner.

Another key hub for advanced data analytics research and education at Purdue is the **Business Information and Analytics Center (BIAC) at the Krannert School of Management**. As its name implies, the Center is focused on IT and analytical techniques and technologies that can be applied to big data in business applications. Like IDSI, it has been structured to facilitate industry-university collaborations, including collaborations with both faculty and students.

The breadth of Purdue's science and engineering capabilities, in combination with University capabilities in data sciences, means that there is a rich and diverse environment of associated applied research at the University. Within the College of Engineering, for example, two "major initiatives" of direct relevance include:

- **Initiative in Data and Engineering Applications.** This multidisciplinary initiative brings together expertise and applied projects in neuromorphic computing, signal processing, data visualization, infrastructure monitoring, geomatics, digital agriculture, advanced manufacturing, biomedical imaging, and education data analytics. Incorporated under the Initiative are subsidiary research centers, including:
  - The **Visual Analytics for Command, Control, and Interoperability Environments Center** (VACCINE) which is a U.S. Department of Homeland Security Center of Excellence focused on the creation of methods and tools, and the application of these methods and tools, to analyze big data in homeland security applications.
  - The previously cited **Center for Brain-Inspired Computing Enabling Autonomous Intelligence**
  - The digital agriculture focused **Open Ag Technology and Systems Group**.

- **Autonomous and Connected Systems (ACS)**, a Purdue College of Engineering initiative focused on advancing the science and engineering of autonomy, robotics, and the Internet of Things (IoT). The Initiative comprises interdisciplinary teams and is partnering with external academic partners, corporations, and international organizations to promote and advance R&D in robotics, and IoT engineering and science.

Beneath the level of the Major Initiatives, the College of Engineering includes a broad variety of other significant research programs and centers focused on data sciences and AI associated applications. Just some examples include:

- **The "Factory of the Future" program.** This is an NSF funded multi-university research center, led by Purdue and also including MIT and IU, focused on development of simulation technologies to model interactions between workers, robots and machines.
- **The Automation and Aging project.** Funded by the NSF, this program is focused on understanding the interactions of different groups of older adults with autonomous systems, and to begin to develop tools that support their effective use.
- **Automatic Control (AC).** Within the School of Electrical and Computer Engineering there has been a significant history of contributions to servomechanisms, manufacturing, automation, and space exploration. AC faculty are now working to address control problems in systems with mobility and autonomy capabilities such as "large robotic swarms and automated transportation systems, smart homes and cities, humanoid robots, and unmanned systems in the air, land, and sea."[42] Work includes innovation in "algorithms for coordination and control of large-scale systems, machine-learning techniques for perception and cognition capabilities that allow the systems to adapt to unstructured/unknown environments, and game-theoretic approaches to ensuring security and reliability."[43]
- **The Intelligent Systems and Assistive Technology (ISAT)** Lab is developing systems for remote surgery. The focus has been on telesurgery technology and approach development that lets highly experienced surgeons remotely assist physicians and less-experienced medical personnel perform complicated procedures in war zones, areas experiencing natural disasters, or underserved and difficult to access rural areas.

The above represent just a sampling of the applied R&D programs taking place at Purdue that integrate advanced data sciences. The IDSI inherently recognizes the diversity of data analytics projects and core competencies across the University and is being developed as the natural entry point for external stakeholders seeking to identify potential collaborators for analytical R&D projects within Purdue's large and diverse research community.

c. Purdue Higher Education

Purdue's Integrative Data Science Initiative (IDSI) is also developing as an organizational entity for helping understand and coordinate the diverse fundamental and applied areas of data sciences educational programs across the University. IDSI is being structured to grow a "Data Science Education Ecosystem" that will prepare "all Purdue students to invent, innovate, and lead in a data driven world." IDSI is working to make this happen not only through traditional coursework and lab activities, but also through "setting up innovative learning communities, undergraduate research opportunities, extracurricular opportunities, distinguished guest speakers, and other events that infuse knowledge, skills and abilities about data science."[44] Specifically, through IDSI coordinated initiatives, Purdue will:

- **Grow and support data sciences related majors.** Purdue currently offers multiple data sciences related majors across campus. IDSI will support and help grow these major areas.
- **Embed data sciences into domain curricula.** Data sciences is viewed as impacting almost every discipline and domain on Purdue's campus. It is expected that students who graduate from Purdue domain majors "will need to

42 https://engineering.purdue.edu/ECE/Research/Areas/AControl
43 Ibid
44 https://www.purdue.edu/data-science/education/index.php

work, lead, and make decisions throughout their career based on data and information from data scientists."[45] An IDSI's goal, therefore, is to "create opportunities for Purdue students from any major to graduate with experience in solving real data science problems relevant to their domain area."[46]

- **Build data literacy & fluency in all students.** "The initiative recognizes that not every student should major in nor seek expert competency in data science; however, ALL Purdue students regardless of major should be data literate. As part of this initiative, Purdue will define data literacy/fluency outcomes for undergraduate students as well as map and make available courses, content and experiences that meet these outcomes."[47]

One of the signature initiatives underway at Purdue for enhancing education for students across multiple disciplines who have an expressed interest in applications of data sciences and advanced analytics to their studies is **"The Data Mine"**. This initiative has established Hillenbrand Hall as a residential learning and research community for student engagement in data sciences (see text box below).

The Data Mine

The Data Mine is a living, learning and research-based community created to introduce students to data science concepts and equip them to create solutions to real-world problems. Members of The Data Mine will be part of a team, living, studying and ultimately, performing data-driven research together. The Data Mine is part of Purdue University's Integrative Data Science Initiative, which is designed to train students across all majors with the data literacy needed to succeed in a data-driven world.

The Data Mine experience allows students the opportunity to live and learn in a community that revolves around a particular topic or theme. The topics and themes studied within The Data Mine incorporate components relevant to any field of study. Students in each community will live in Hillenbrand Hall, where they will have access to renovated study space, upgraded learning technology and hands-on interaction with faculty within the hall.

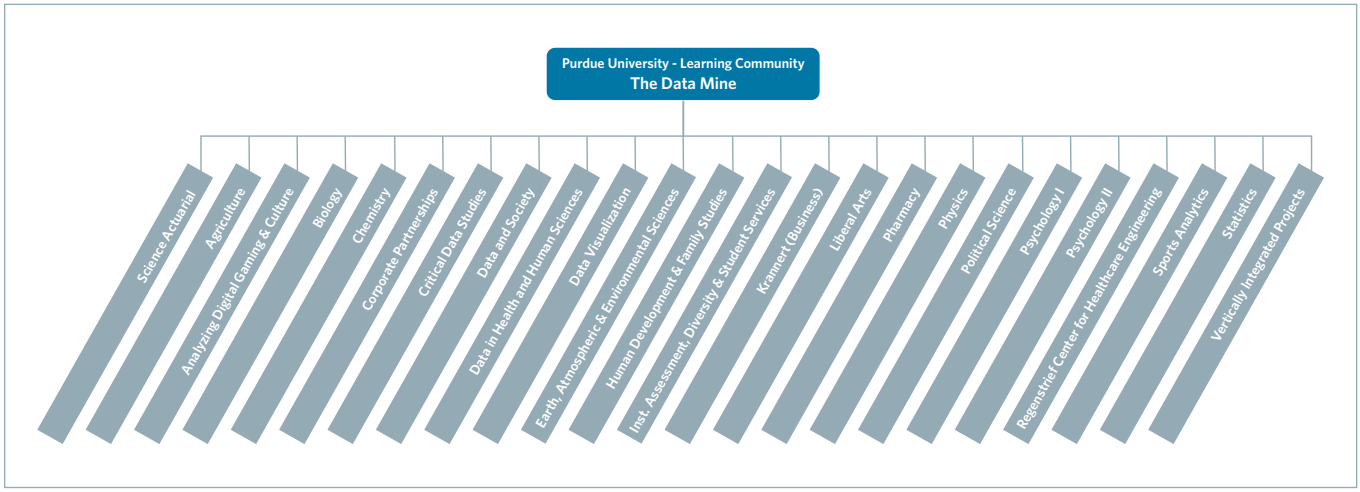
The Data Mine is open to students from any major of study. Students will learn some of the skills most sought after by companies and graduate programs. No computational background is required. The key trait for joining The Data Mine is the desire to learn data science in a rigorous, but welcoming environment.

https://www.purdue.edu/learningcommunities/profiles/data_mine/index.html

The Data Mine represents an ambitious effort to build a large-scale cohort of students at Purdue with fundamental expertise in data sciences theory, tools and techniques and their application to particular domain areas. Already operational with its first intake of students, the breadth of domain areas set to be covered by The Data Mine in relation to the application of data sciences shows how ambitious this project is (Figure 11). As Figure 11 illustrates, cohorts span the liberal arts, humanities, social sciences, business, engineering and scientific disciplines.

45 *Ibid*
46 *Ibid*
47 *Ibid*

Figure 11: "The Data Mine" Learning Community at Purdue University



In addition, of course, to The Data Mine focus areas, Purdue also has a series of undergraduate and graduate degree and certificate programs specifically focused on advanced analytics and data sciences. Formal programs of study include, for example:

- BS in Data Visualization,
- BS in Data sciences (a collaboration between the Departments of Computer Science and Statistics),
- MBA with Specialization in Business Analytics,
- MS in Business Analytics and Information Management,
- MS in Geodata Science, and
- Certificate in Applied Statistics.

Purdue also sustains an intensive commitment to the development of online courses of study, and the development of customized certification and continuing education study for industry and other external audiences.

Purdue's degree and non-degree award totals in core advanced analytics-related fields totaled 1,353 in 2017 (Table 14). Nearly 1,100 graduates were at the Bachelor's level with leading totals in a varied set of key fields spanning computer science and IT, math and statistics, computer graphics, economics, and computer engineering. Purdue is generating significant numbers of computer science and IT graduates at the Master's and Doctoral levels.

Table 14: Purdue (All Campuses) Graduate Totals in Degree Fields Most Closely Aligned with Advanced Data Analytics, by Degree Level, 2017

Degree Field	Certificates & Awards	Bachelor's degree	Master's degree	Doctorate	Total
Computer Science		240	77	28	**345**
Mathematical Statistics and Probability		199			**199**
Information Technology	1	146	49		**196**
Computer Graphics	1	138	19		**158**
Economics, General		118	29	8	**155**
Computer Engineering, General		115			**115**
Mathematics, General		79	14	13	**106**
Statistics, General	6		18	9	**33**
Mathematics and Computer Science		12			**12**
Applied Mathematics, General		11			**11**
Financial Mathematics		8			**8**
Operations Management and Supervision	7				**7**
Management Information Systems, General	1	6			**6**
Computational and Applied Mathematics		1			**1**
Management Information Systems and Services, Other					**1**
Total	**16**	**1,073**	**206**	**58**	**1,353**

Source: National Center for Education Statistics, IPEDs postsecondary degree database.

d. Purdue Supporting Infrastructure

Purdue has extensive computational infrastructure for supporting fundamental and advanced research in advanced analytics, AI and associated data intensive fields. Operated under Information Technology at Purdue (ITaP), the University's high performance computing and associated data storage infrastructure is primarily organized around shared research computing clusters. Some of the key computational cluster assets available to facilitate basic through applied data sciences work are:

- "Weber", which is a new specialty Community Cluster for restricted and controlled data (serving applications and research that are restricted by ITAR or DFARS 800-171 controls). Developed in August 2019, Weber currently consists of HP compute nodes with two 10-core Intel Xeon-E5 "Haswell" processors (20 cores per node) and 64 GB of memory. Purdue also has the "REED GovCloud" which is also designed for working with data encumbered with federal security regulations.
- "Gilbreth" is a Community Cluster designed for applications which are able to take advantage of GPU accelerators.

- "Scholar" is a cluster open to Purdue instructors across the entire university in any field or discipline that requires access to supercomputing class computational capacity.
- "Data Workbench" provides a resource for non-batch big data analysis and simulation.
- "Brown", "Halstead" and 'Rice" are three Community Clusters optimized primarily for research communities running traditional, tightly-coupled science and engineering applications.
- "Snyder" is a Community Cluster which is optimized for data intensive applications requiring large amounts of shared memory per node, such as life sciences.
- "Hammer" is focused towards Purdue research communities that utilize loosely-coupled, high-throughput computing.

Purdue also operates a Bioinformatics Core to provide data analysis services for Next Generation Sequencing data in biological research.

Because of the importance of data visualization, Purdue has the Envision Center which works with individual researchers and research teams to help identify best pathways to data visualization and communication of results via visualization applications. The Envision Center also provides development support for virtual environments for training and simulation.

Given the size of the University and its varied computational and analytics systems, the University operates a "DiaGrid Hub" to connect users to the diverse set of advanced computing resources and help the Purdue science community share resources and applications. DiaGrid is set-up to support scalable access to University computation resources through online tools, and is self-service – allowing users to upload their own content – including scientific applications and tools, datasets, tutorials, courses, publications, etc., and share with other users, groups or publicly. As noted by ITaP:

> *Users may launch simulations and analyze results via an ordinary web browser without downloading, compiling, or installing any software. Tool developers can follow an automated process to create, install, test, and publish interactive tools on the DiaGrid web site. Computational tools on DiaGrid include bioinformatics (BLAST, BEAST, etc), simulations (GROMACS, NAMD, CryoEM, etc), visualization tools (ParaView, PyMol, etc) and general purpose computational tools such as IDEs with plugins for submitting jobs to HPC resources (e.g., RStudio, Spyder for Python, SubmitR for serial, parallel and parameter sweeps jobs, etc). DiaGrid.org also hosts science gateways for research groups.*[48]

48 *https://www.rcac.purdue.edu/services/diagrid/*

3. University of Notre Dame

The University of Notre Dame has achieved substantial growth in research activity. As noted by TEConomy in the October 2018 report on funding at Indiana research universities:

> *The University of Notre Dame is a special case. Traditionally a teaching-focused institution, Notre Dame has been managing a deliberately planned transition to become a top-tier research university. The impact of this strategy is showing up in the growth in overall research funding, including external funding. The average research university in the United States has research expenditures of $112.2 million in 2016, and Notre Dame now stands at close to double that with $202.2 million in 2016 research expenditures. Notre Dame is now in the elite group of top-quartile research expenditure institutions. Achieving this level of performance has required significant investment of internal university funds to support the development of faculty, infrastructure, and supporting resources that allow the university to compete successfully for external research funds. The investment of the university's own resources in building research stature is paying off in terms of Notre Dame's growing research portfolio and structure—and is partly explanatory for the higher-than-average institutional funding profile within Indiana.*[49]

Notre Dame current has a total student population of 12,607 students (8,617 undergraduate students and 3,990 graduate students) and 1,309 faculty.[50]

The University has been proactively growing its research volume and research profile by strategically focusing on seven strategic university wide research areas and ten cluster hiring opportunities identified as being areas of historic research core competency and relevance to future opportunities. These areas have been supplemented by research foci of the colleges. These areas have been designated by the University for further priority investment. Data analytics will be of cross-cutting relevance and importance to most. Two of the cluster-hiring themes are explicitly oriented to advancing data sciences and advanced data analytics methods and applications, these being Applied and Computational Mathematics and Statistics, and Computational Data Science and Engineering. In addition, Notre Dame is expanding its long-standing history of contributions in the field of ethics into the important space of ethics in relation to artificial intelligence and machine learning. University leadership notes that this will be a significant area of emphasis bringing together existing faculty and recruitment into multiple new faculty positions (see text box "Ethics and Artificial Intelligence: A Developing Indiana Specialization?").

a. University of Notre Dame Fundamental Data Science Development

The Applied and Computational Mathematics and Statistics (ACMS) initiative at Notre Dame is engaged in fundamental mathematical and statistical development work for advanced analytics and, as its name implies, also contributing to the application of advanced mathematics and statistics to multi-disciplinary research focus areas across the university. In fundamental research ACMS is engaged in the development of models and tools for Big Data analytics, deep learning in complex data, and AI methods. ACMS is also working on novel mathematical and computational tools for network science.

The Department of Computer Science and Engineering (CSE) is engaged in fundamental research in computational methods and application of computational and advanced analytics methods across a range of applied needs and challenges. Work ranges from basic inquiry (such as the development of novel computational tools for network analysis) through to applications (such as the development of an AI enabled virtual assistant for use by software engineers).

The Computational Data Science and Engineering (CDSE) initiative is housed jointly within the Notre Dame Department of Aerospace and Mechanical Engineering and the Department of Computer Science and Engineering.

49 *Martin Grueber, Ryan Helwig and Simon Tripp. "Assessing R&D Funding Across Indiana's Major Research Universities." October 2018. TEConomy Partners, LLC. Prepared for BioCrossroads.*

50 *https://www.nd.edu/about/*

Fundamental research activity is taking place in the development of computational and mathematical tools for multi scale predictive computations. The research team is also engaged in studies of how "data science, machine learning, and artificial intelligence can accelerate engineering science and provide understanding of the uncertainties in a simulation."[51] Research is being undertaken in areas that "impacts materials science, combustion, fluid mechanics, high-energy density physics, and biology."[52]

As a joint initiative, the College of Engineering and College of Science jointly operate the **Center for Informatics and Computational Science (CICS)** which is "grounded in the mathematical, statistical, and scientific computing disciplines and aims to address unifying themes to allow predictive modeling of complex scientific and engineered systems."[53] CICS is conducting fundamental research inquiry in four thematic areas:

- Computational Mathematics
- Computational Statistics
- Machine Learning
- Scientific Computing.

51 https://ame.nd.edu/research/computational-engineering
52 Ibid
53 https://cics.nd.edu/about/

Ethics and Artificial Intelligence: A Developing Indiana Specialization?

Ethics may be defined as "the moral principles governing the behavior or actions of an individual or a group." As humans we have developed a moral code of ethics that generally defines what we consider to be right, just or fair. In many case these ethics are coded into the laws that govern our society and the allowable actions of individuals within it. Individuals make decisions, and those decisions may be determined to be ethical or unethical. With the rise of "machine intelligence" humankind is entering a new era where artificial systems will be making the equivalent of decisions. This has given rise to a field of machine ethics, which is concerned with the moral behavior of artificial agents and systems (which may be cyber-physical, e.g. an autonomous robot, or purely cyber, e.g. a machine learning software algorithm).

The new age of intelligent machines brings with it a broad series of ethical questions, challenges and potential issues.

[54]Nalini notes that there is a landscape of issues collectively contained under the "Ethics of AI". These comprise issues that point to either the ethical quality of the predictions or of end outcomes or their direct and indirect impact to humans. Nalini summarizes these as:

In the realm of "what AI is" (i.e. datasets, models and predictions):

1. *Bias and Fairness*
2. *Accountability and Remediability*
3. *Transparency, Interpretability and Explainability*

In the realm of "what AI does" are issues of:

1. *Safety*
2. *Human-AI interaction*
3. *Cyber-security and Malicious Use*
4. *Privacy, Control and Agency (or lack thereof, i.e. Surveillance)*

In the realm of "what AI impacts" are issues related to:

1. *Automation, Job loss, Labor trends*
2. *Impact to Democracy and Civil rights*
3. *Human-Human interaction*

In the realm of "what AI can be" are issues related to threats from human-like cognitive abilities and concerns around singularity, control going all the way up to debates around robot rights (akin to human rights).

As advanced analytics and AI are deployed across commerce and society, the above ethical considerations are critically important to address. This presents an opportunity for research in the field and also for consultative services, testing services, etc. for businesses, government and other organizations deploying AI and machine intelligence systems. It may be an opportunity for Indiana.

Indiana University, Purdue University and the University of Notre Dame are each engaged, or engaging, in well-funded initiatives focused on the ethics of AI. In the case of Indiana University and Purdue University major funding has been provided by Lilly Endowment. Purdue has received a $491,000 planning grant from Lilly Endowment to support an initiative called "Leading Ethically in the Age of AI and Big Data" which is a multidisciplinary program bringing together experts in "academia, business, government and other fields to help foster character and ethical values in their students as they learn the science and technology relating to digital technologies, such as artificial intelligence and Big data management."[55] At Indiana University, Lilly Endowment has provided a planning grant of $349,000 to support development of "Developing Character for a Digital World", which is intended to "bring together faculty and

54 B Nalini. "The Hitchhiker's Guide to AI Ethics. A 3-part series exploring ethics issues in Artificial Intelligence." Medium. Towards Data Science. Accessed online at: https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0

55 Purdue University News. "Lilly Endowment grant to help Purdue address the intersections of ethics and technology." September 25, 2019. Accessed online at: https://www.purdue.edu/newsroom/releases/2019/Q3/lilly-endowment-grant-to-help-purdue-address-the-intersections-of-ethics-and-technology.html

outside experts to develop an integrated curriculum that prepares faculty and students to confront current and future ethical challenges of our increasingly digital world."[56] The University of Notre Dame is similarly engaging in a major transdisciplinary initiative to study ethics and AI. Notre Dame's is building upon a series of major conference events held at the University (including the Mendoza College of Business "Artificial Intelligence and Business Ethics: Friends or Foes?" event held in 2018, and "The Future of Work" 2019 Conference which focused on AI and automation impacts on future human labor demand) to put together a major university-wide Technology and Ethics initiative that will have as a primary thrust areas AI and Ethics. Notre Dame has been actively hiring new faculty, with more to come over the next five years, forming an intellectual core to work on both fundamental data science questions and ethics, with a goal of training data scientists with a deep understanding of ethics. This capacity is being built with significant contributions from benefaction as one of the priority fundraising areas of the campus and with connectivity to major Notre Dame alumni-led corporations. It is anticipated that the Notre Dame initiative will also build on the University's existing research competency in ethics embodied in the "de Nicola Center for Ethics and Culture" which seeks to apply the Catholic moral and intellectual tradition to ethical questions and challenges through teaching, research, and dialogue.

56 Ibid.

b. University of Notre Dame Data Science Application Areas

As noted previously, quantitative methods and advanced data analytics form a cross-cutting competency that is applied to a broad variety of research areas across the University. Some of the areas of emphasis that are most of relevance to advanced industry clusters observed in Indiana include applied data sciences work focused in biological science and in engineering applications.

The Applied and Computational Mathematics and Statistics (ACMS) initiative, as its name implies, is focused principally on porting the skills and expertise of Notre Dame faculty in mathematics and statistics to work collaboratively on multidisciplinary projects with faculty from diverse domain fields. ACMS faculty are working together with biological scientists, chemists, economists and econometricians, engineers, political scientists, psychologists, and sociologists, for example. As noted by the program:

> *ACMS provides the expertise in mathematical and computational modeling and statistical analysis necessary to the University's contribution to some of today's most prominent research areas. The Department of Applied and Computational Mathematics and Statistics aspires to create an environment for excellence in which faculty and students who have deep intellectual expertise can transcend disciplinary boundaries to impact critical problems in the natural and social sciences, technology, and beyond. The department's outward focus will uncover new areas in which applied and computational mathematics and statistics inform the discussion of a complex issue and find common patterns in diverse subjects, all of which enable the fruitful cross-fertilization of ideas.*[57]

Areas at ACMS where there are clusters of multiple ACMS faculty engaged include:

- **Mathematical and Computational Biology** — Multiscale modeling, using a combination of discrete stochastic systems and differential equations, of biomedical problems including blood clot formation, spread of infection, development, and cancer.
- **Numerical Differential Equations** — The design, efficient implementation, and analysis of numerical methods for solving differential equations arising in science and engineering.
- **Numerical Algebraic Geometry** — The discovery, implementation, and application of algorithms to numerically compute and manipulate the solution sets of systems of polynomials.

57 https://acms.nd.edu/about/

- **Bioinformatics and Biostatistics** — The application of statistical and computational methods to biological and medical data to model, analyze, and predict biological processes.
- **Applied Partial Differential Equations** — Modeling and analysis using partial differential equations tools and theories to study real-world problems arising from the natural and social sciences and engineering.
- **Scientific Computing** — The construction and implementation of mathematical algorithms to run on large parallel high-performance computers and their application to problems in science, engineering, and social science.

The Department of Computer Science and Engineering (CSE) has active programs in applied research using AI, machine learning, network analytics, natural language processing and machine vision. Areas evident in CSE's advanced analytics and AI oriented work include:

- **Computational Biomedicine and Health**, including a focus on biomedical imaging problems using machine learning, data mining, optimization, and geometric computing techniques.
- **Computer Vision Research Laboratory (CVRL)** conducting work in AI-based visual analytics systems and applications of biometrics, machine learning, biologically-inspired algorithms, media forensics, and the digital humanities.
- **Data Mining toward Decision Making (DM2)** Lab developing data-driven computational methods and intelligent systems to facilitate human decision making.
- **Visualization Lab** developing algorithms and techniques for data visualization.

The Center for Research Computing (CRC) is discussed below under "Supporting Infrastructure", but also operates as a multidisciplinary research environment that supports collaborations and applied innovation using advanced computation, software engineering, data analysis, and other digital research tools. The Center enhances the University's innovative applications of cyberinfrastructure, provides support for interdisciplinary research and education, and conducts computational research.

Given Notre Dame's history of work in social science disciplines, the University also operates the **Center for Social Science Research (CSSR)** as an organization supporting the application of advanced analytics in social sciences research. CSSR provides expertise in in research design, statistics, data collection, modeling and analysis, geospatial (GIS) analysis, and visualization.

Another major initiative is **iCeNSA** which brings interdisciplinary teams together in a center that is focused around applied network and data science problems in "social, biological, physical, environmental, financial, business, and defense systems."[58] As a research center, iCeNSA notes that it focuses on:

- Developing a systems-level understanding of the fundamental processes and mechanisms that underly the structural, dynamical and functional properties of complex systems.
- Developing and integrating novel mathematical and computational algorithms and tools for network and sciences.

iCeNSA also serves an education mission, providing a "hands-on training environment in network science for undergraduates, graduate students and postdoctoral associates" responding to a "growing need in academia, industry and government for personnel who have a problem solving mentality, an interdisciplinary and global understanding of networked systems and the ability to design practical and sound solutions using rigorous mathematical and computational tools."[59]

58 https://icensa.com/content/about
59 Ibid

Computational Engineering is also a focus area at Notre Dame that applies advanced analytics and high-performance computing capacity to applied engineering challenges. Research in this area:

> *Spans across multiple domains to apply cutting edge mathematics, computer science, and statistics to utilize the fastest computers in the world and the next generation of high-performance computing systems... The Computational Engineering research group also studies how data science, machine learning, and artificial intelligence can accelerate engineering science and provide understanding of the uncertainties in a simulation.*[60]

Application areas for faculty research include optimization, inverse problems, uncertainty quantification, high-fidelity simulation, reduced-order modeling, and scientific machine learning.

The Notre Dame **Center for Informatics and Computational Science (CICS)** is also applying advanced analytics tools and technologies, such as AI and ML, to application spaces, including:

- Hazard Resilience of Civil Infrastructure
- Predictive Materials Modeling
- Ecological Forecasting
- Fluid Mechanics
- Epidemiology
- Monitoring and Maintenance of Civil Infrastructure
- Terrestrial Ecosystems and Climate
- Cardio/Cerebrovascular Modeling
- Uncertainty Visualization.

c. Higher Education at the University of Notre Dame

The University of Notre Dame has multiple undergraduate and graduate programs that focus on data sciences or include a major component of the curriculum focused around advanced analytics and applications. Chief among the graduate programs are:

Table 15: Degree (Undergraduate and Graduate) and Certificate Programs at Notre Dame in Advanced Analytics and Associated Fields.

Notre Dame College, School or Department	Degree/Qualification Programs
Department of Applied and Computational Mathematics	• BS in Applied and Computational Mathematics and Statistics.[61] • Professional MS in Applied and Computational Mathematics and Statistics (MSP-ACMS) – (a terminal degree). • MS in Applied and Computational Mathematics and Statistics (research oriented Masters for doctoral students). • PhD in Applied and Computational Mathematics
Online	• MS in Data sciences (offered online with immersions at Notre Dame and Silicon Valley sites).[62]

60 *https://ame.nd.edu/research/computational-engineering*

61 *In addition to the core BS major, ACMS also offers a concentration in biological sciences, which "prepares students for further study or employment in computational biology, bioinformatics, ecological modeling, or epidemiology, as well as an ACMS supplementary major."*

62 *The online MS in Data Science has been developed in collaboration with AT&T, which then brings actual industry challenges for integration into course work.*

Notre Dame College, School or Department	Degree/Qualification Programs
Mendoza College of Business	• BS in Business Analytics. Located in the Department of Information Technology, Analytics & Operations (ITAO). • MS in Business Analytics (MSBA). Covers: data mining; data storage and manipulation; data visualization, statistics, modeling, optimization and simulation.
College of Engineering	• Computer Science BS, MS and PhD degrees, including analytics focused degrees through the Department of Computer Science and Engineering • Minor in computational engineering in AME
College of Arts and Letters	• Data Science minor

Notre Dame produced 534 graduates in degree fields most closely aligned with advanced data analytics in 2017 (Table 16). The degrees span a varied set of fields including economics, several areas in math and statistics, MIS, and computer sciences. In the School of Engineering, computer engineering stands out in its generation of graduate level degrees.

Table 16: Notre Dame Graduate Totals in Degree Fields Most Closely Aligned with Advanced Data Analytics, by Degree Level, 2017

Degree Field	Bachelor's degree	Master's degree	Doctorate	Total
Economics, General	146	1	2	**149**
Management Information Systems, General	92			**92**
Mathematics and Statistics, Other	66	15	4	**85**
Computer and Information Sciences, General	81	14	13	**108**
Mathematics, General	54	14	5	**73**
Computer Engineering, General	20			**20**
Mathematics and Statistics	6			**6**
Computer and Information Sciences and Support Services, Other	1			**1**
Total	**466**	**44**	**24**	**534**

Source: National Center for Education Statistics, IPEDs postsecondary degree database.

d. Supporting Infrastructure

As Notre Dame's research intensity has grown, so too has its investment in high performance computing infrastructure. The Center for Research Computing is the lead organizing entity for assuring the University has access to state of the art computing and communication infrastructure, resources, and skilled computing staff. CRC personnel are also available to assist researchers in examining potential applications of high-performance computing to specific research needs and provides access to training and specific hardware and software tools and University resources.

The Center for Research Computing provides access to several major computer cluster resources and operates the University's supercomputing infrastructure. The CRC provides access to grid networks and storage resources, while

also providing user support services and expertise to the internal Notre Dame community and also to external organizations and industry. The CRC is comprised of four main groups with complementary expertise:

- Computational Scientists. Function as primary initiators of collaborations between CRC groups and external researchers. Able to operate as as principal investigators together with faculty and industry partners.
- Software Development. Software development provides "agile, security-oriented, programming teams that empower research through innovative solutions for sharing, processing, analyzing, and visualizing data."[63]
- Center for Social Science Research. Providing expertise in research design, data collection, modeling analysis, geospatial (GIS) analysis, and visualization. This is a function of the Center for Social Science Research (CSSR).
- High Performance Computing. Providing access to high performance computing and user support for associated hardware and software.

The CRC also incorporates expertise in cybersecurity, visualization, and embedded systems, as well as a business office.

Conclusions from the University Review

It is evident that the state's research universities all have significant investments and growth in foundational data sciences and applied analytics capabilities that support talent pipelines and the development of applied analytical skills.

- **Interdisciplinary institutes and programs in advanced data sciences**
  - Purdue Integrative Data Science Initiative (IDSI); Notre Dame Center of Network and Data Sciences; IU School of Informatics, Computing, and Engineering.

- **Business analytics programs** with significant graduate volume and major corporate partners
  - Analytics-focused programs and engaged corporate partners at Mendoza College of Business, Krenicki Center for Business Analytics & Machine Learning, Kelley School of Business.

- **Data sciences training and immersion programs** that expose students to real-world data science problems and workflows, examples of which include:
  - Data Mine at Purdue – signature example of industry-facing, immersive talent pipeline program.
  - Online Masters for Data sciences program at Notre Dame – intensive mid-career up-skilling.
  - Modular certification and courses offered through IU Kelley School – exposure to corporate partners.

There are a broad variety of data sciences and advanced analytics programs available in Indiana and more in development (on campus, online, full or part-time). Table 17 lists many of these degree and certificate programs, and TEConomy's analysis of NCES IPED data indicate there have been 3,926 graduates in programs relevant to data sciences, advanced analytics and AI in 2017 at the three Indiana research universities.

63 https://crc.nd.edu/about/

Table 17: Indicative Summary of Programs of Study at Indiana's Universities in Data sciences, Advanced Analytics and AI.

Institution	Examples of Programs
IU Bloomington	BS and MS programs in Data sciences; PhD Minor in Data sciences; MS in Business Analytics; Online Certificate in Data sciences; Business Analytics Certificate Program; MBA with Major in Business Analytics.
IU or Purdue Degrees and Certificates at IUPUI	MS or PhD Minor in Health Informatics; MS in Bioinformatics; MS in Informatics (Data Analysis Specialization); PhD in Data sciences; PhD in Health and Biomedical Informatics; MS in Mathematics and Applied Statistics; Online Clinical Informatics Certificate; Health Information Management and Exchange Certificate.
Purdue	Integrating data sciences across the university (cross-cutting). Data Mine living and learning community. Data Mine program for all engineering students. BS in Data Visualization; BS in Data sciences; MBA with Specialization in Business Analytics; MS in Business Analytics and Information Management; MS in Geodata Science; Certificate in Applied Statistics.
Notre Dame	MS in Data Science; MS in Business Analytics; MS in Applied Statistics; MS in Applied and Computational Mathematics-Statistics; Master's Study in Predictive Analytics.
Rose-Hulman	BS in Biomathematics; Minor in Data sciences
Valparaiso U.	MS in Analytics and Modeling
Ball State	BS in Business Analytics; Graduate Certificate in Statistical Modeling

C. Talent Flow, Availability and Access in Indiana

As established in previous sections, the most important need for any AI-enabling ecosystem is a pipeline of skilled talent that can utilize the tools and methods in this space towards applications. For Indiana, this need was consistently cited in industry interviews as one of the most important determinants of success for companies that have a desire to build out their capabilities in AI-related technologies and services.

Given the cross-disciplinary nature of skills and business experience encapsulated by the AI space, it is difficult to partition a definitive subset of employment working specifically in this technology area. Moreover, the importance of embedding knowledge about business models and markets into development of AI-driven solutions for companies means that some leading employment segments that are active in this area may not be aligned with traditional computer science or statistical backgrounds. However, focusing on a broader "AI-enabling" labor force that contributes key skill sets that can be leveraged by industry to drive applied solutions as a part of data sciences and analytics teams can still be informative to understanding Indiana's current talent profile.

For the purposes of analyzing trends for Indiana, the AI-enabling workforce can be defined across a spectrum of different occupational categories spanning technical computer science positions to business support functions. Table 18 below shows these categories and the associated 2018 Indiana employment.

Table 18: AI-Enabling Occupation Segments and 2018 Indiana Employment

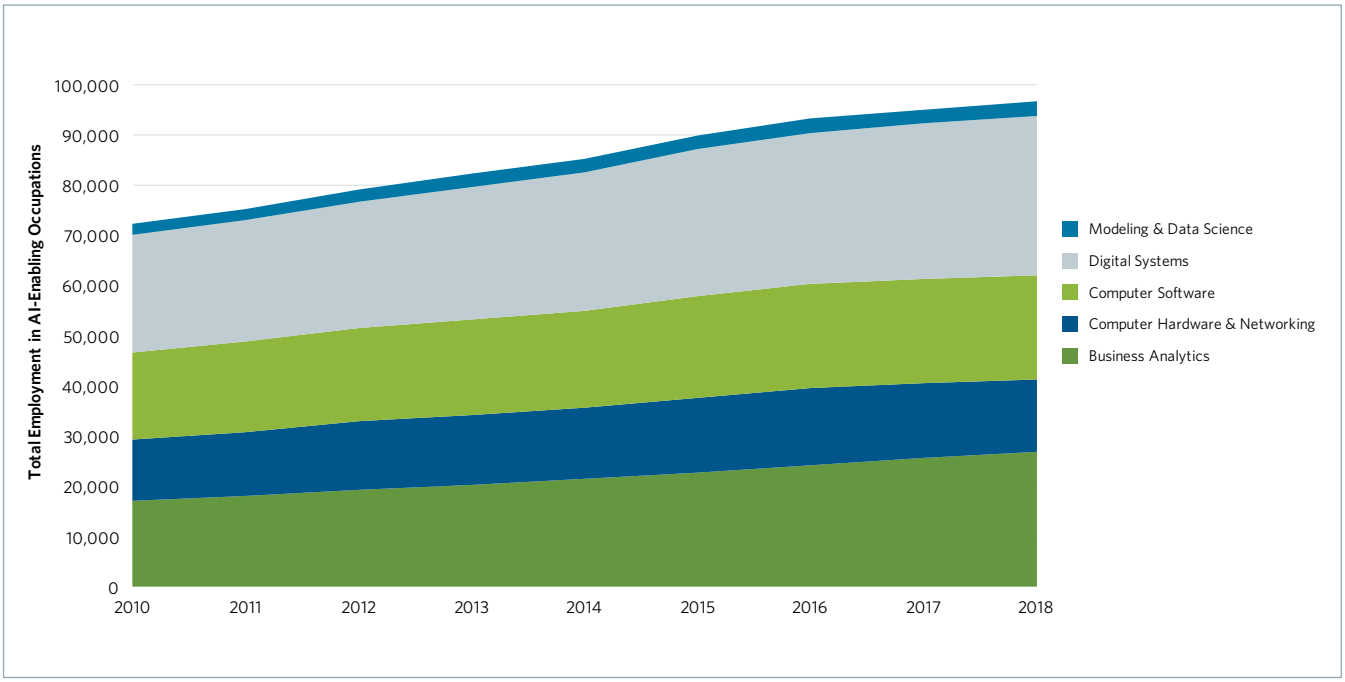
AI-Enabling Occupation Category	Occupation	2018 Indiana Employment	% Change in Employment, 2010-2018
Business Analytics	Management Analysts	10,198	32.1%
	Market Research Analysts and Marketing Specialists	9,507	121.0%
	Financial Analysts	3,176	36.9%
	Logisticians	2,742	70.5%
	Credit Analysts	718	10.1%
	Budget Analysts	538	-16.2%
Computer Hardware & Networking	Network and Computer Systems Administrators	6,623	13.3%
	Computer Network Architects	3,500	52.4%
	Computer Network Support Specialists	1,878	22.6%
	Information Security Analysts	1,101	113.8%
	Computer Operators	879	-39.1%
	Computer Hardware Engineers	424	-17.7%
Computer Software	Software Developers, Applications	10,003	78.4%
	Computer Programmers	5,047	-25.8%
	Software Developers, Systems Software	4,004	3.1%
	Web Developers	1,641	71.9%
Digital Systems	Computer User Support Specialists	12,055	40.3%
	Computer Systems Analysts	9,787	23.2%
	Computer and Information Systems Managers	5,030	33.3%
	Computer Occupations, All Other	3,074	104.0%
	Database Administrators	1,572	10.8%
	Computer and Information Research Scientists	276	11.3%
Modeling & Analytics	Operations Research Analysts	1,215	74.2%
	Statistical Assistants	690	-4.0%
	Statisticians	556	92.1%
	Actuaries	265	33.8%
	Economists	130	49.4%
	Survey Researchers	127	-38.6%
	Mathematicians	16	-21.4%
	Miscellaneous Mathematical Science Occupations	16	-20.2%
Total AI-Enabling Occupational Employment	96,788	33.8%	

Source: EMSI Occupational Data, EMSI 2019.4

Indiana's total footprint for AI-enabling employment stands at just over 96,788 jobs in 2018, representing approximately 3 percent of the state's total workforce. The workforce is currently dominated by occupations related to the business analytics and digital systems services, which represent 27.8 percent and 32.8 percent of total AI-enabling employment in the state, respectively. As shown in Figure 12 below, workforce levels in these occupations have grown

steadily since 2010, showing gains of 33.8 percent in total employment by 2018. This lags slightly behind overall U.S. growth in these occupational segments of 36.5 percent over the same time period.

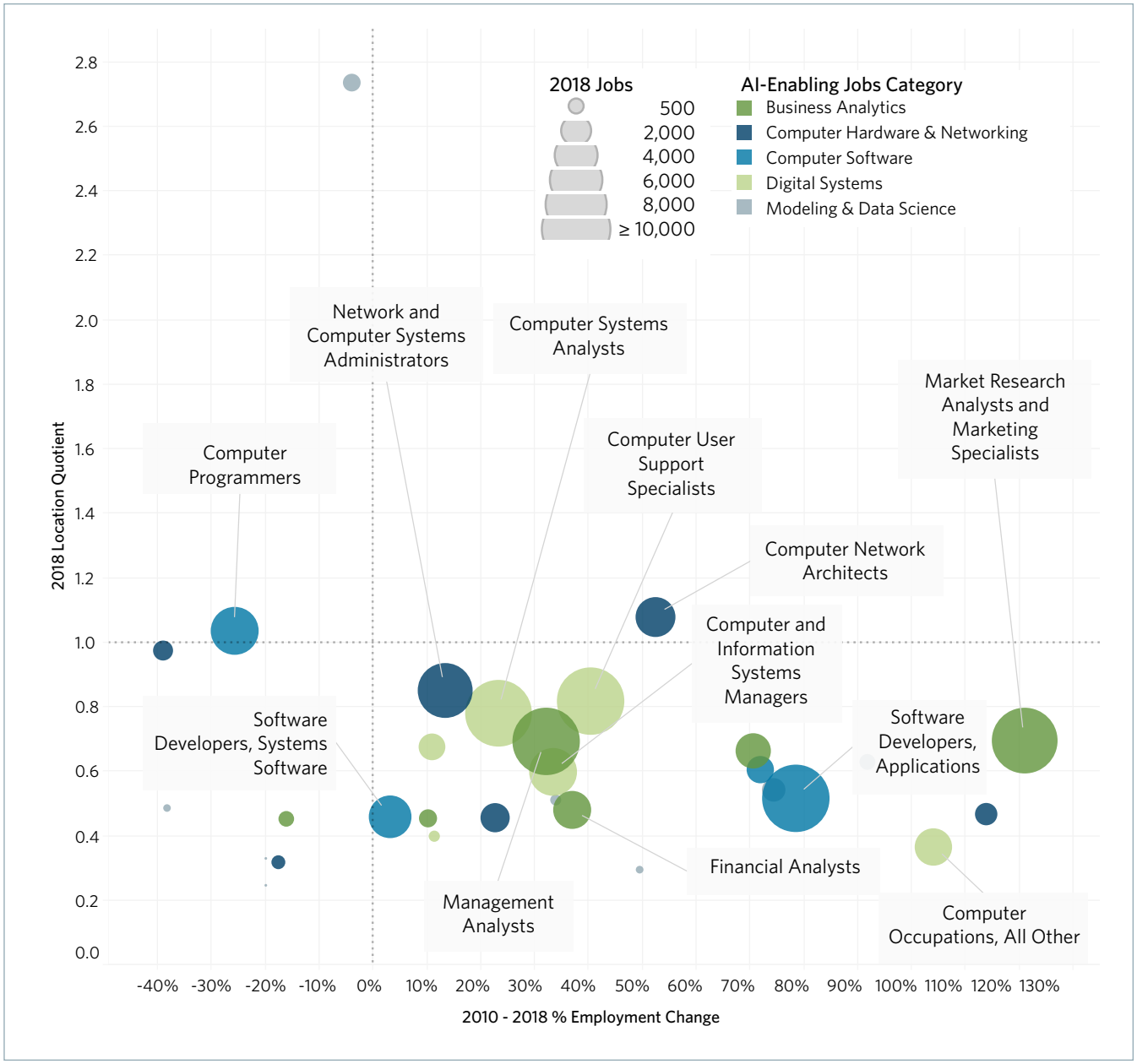
Figure 12: Employment Growth in AI-Enabling Occupational Segments in Indiana, 2010-2018



Source: EMSI Occupational Data, EMSI 2019.4

Despite positive growth, Indiana's position in AI-enabling workforce segments relative to national trends is less robust when considering the need for developing specialized talent bases that are indicative of competitive advantages in attraction and retention. Figure 13 below shows the profile of various AI-enabling occupations. These occupations are displayed in the figure in terms of their employment growth from 2010 to 2018 as well as their 2018 occupational location quotient, a measure of how concentrated or "specialized" the state's employment is in a given occupation relative to the trends in the national labor market (with a value > 1.0 indicating Indiana's occupational employment is more specialized than the national workforce and a value < 1.0 indicating it is not as specialized). Specific occupations are colored based on the overall AI-enabling segment they belong to, with the relative size of each data point indicating the employment footprint in 2018. Occupations that represent especially notable competitive advantages for the state will have both a high growth rate indicating an expanding and healthy base of industry employers as well as specializations greater than national levels that indicate agglomeration of talent within the state relative to the country as a whole.

Figure 13: Detailed Profile of AI-Enabling Occupational Segments in Indiana, 2010-2018



Note: occupations with >3,000 Indiana employees in 2018 annotated.

Source: EMSI Occupational Data, EMSI 2019.4

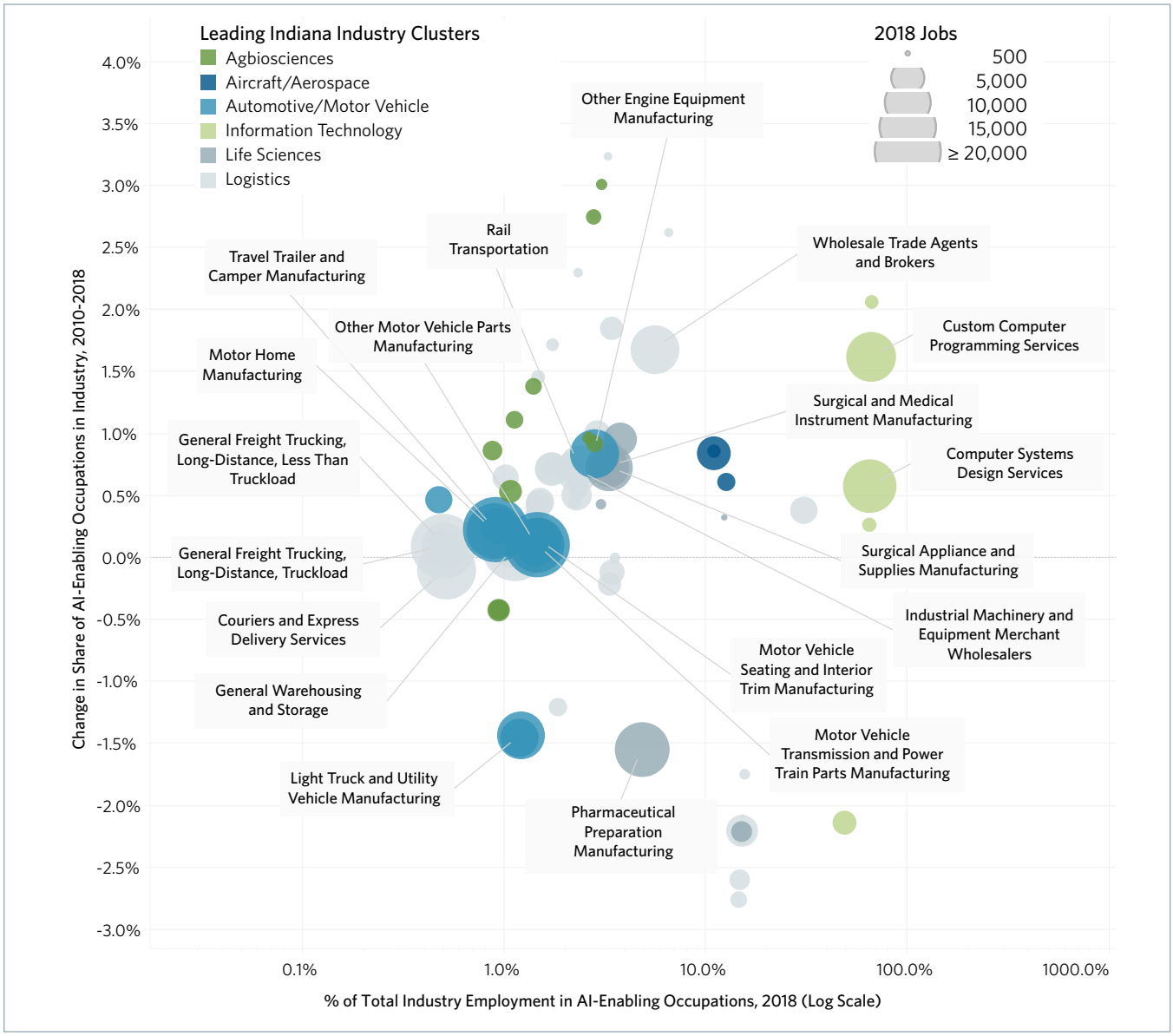
As an illustrative example indicative of the position of many of Indiana's current AI-enabling occupations, consider the detailed occupation of applications software developers. This occupation has a significant 2018 employment footprint in the state of 10,003 jobs and has had significant growth since 2010 of 78 percent. However, relative to the makeup of the overall U.S. workforce this occupation has a specialization of 0.52, indicating applications software developers are nearly half as concentrated in Indiana's workforce as they are nationally. This position reflects the narrative across most other AI-enabling occupations in the state - **although the majority of occupations in AI-enabling segments have experienced significant growth in Indiana over the last nine years, very few are specialized, indicating that Indiana is lagging behind (i.e., is undersized relative to) other regions of the country in building the talent base required to create competitive advantages.**

Another important perspective is the level of utilization of various AI-enabling occupations within industry segments, which can indicate the degree to which industries are incorporating advanced analytics technologies into their business models. Shifts in the proportion of total jobs in an industry cluster that these occupations make up can also indicate recent trends towards upgrading skilled labor forces or shifting the labor mix to meet new demand for advanced technologies. To assess the current position of Indiana's leading industries, Figure 14 shows individual industries at the six-digit NAICS level that aligned with Indiana's leading industry clusters with respect to the percentage of total 2018 industry employment in AI-enabling occupations as well as the change in percentage points (or share) from 2010-2018. Industries that currently utilize high levels of AI-enabling workers will have higher shares of AI-enabling occupational employment, while industries that have made significant efforts to integrate new AI-related positions into their labor mix over the 2010-to-2018 period will have higher shifts in the share of AI-enabling occupations.

The majority of Indiana's industry clusters have lower levels of AI-enabling occupational employment relative to the national average of 4.3 percent, on average totaling around 2.6 percent of the overall workforce within most industry clusters. However, almost 25 percent of the state's employment is contained in industries that have a significant concentration of over 5 percent of occupational employment in AI-enabling workforce segments, making them significant drivers for the state's skilled talent in this space. As shown in Figure 14, most of the state's industry sectors associated with leading traded industry clusters employing above average levels of AI-enabling occupational segments are in information technology, with the highest levels of concentration of these segments located in computer services industries. Other non-traded and business services clusters not shown below, such as consulting and marketing and financial services, also have significant concentrations, with some additional presence in communications, corporate headquarters, research organizations, and publishing. While not as concentrated, other key clusters with above average levels of occupational employment in AI-enabling job segments include engineering services, aircraft and electronics manufacturing, and biopharmaceutical, diagnostic, and lab equipment manufacturing. These industry clusters all require increasingly analytics-driven approaches to business operations and product development, and all leverage skilled talent that is best able to integrate new technologies into business practices.

As industry clusters adapt to analytics-driven business models, products, and services that leverage AI-enabled intelligence, it is expected that the mix of skill sets employed by traditional industries will also change to align with the new environment. As noted above, Figure 14 also incorporates one possible indicator of shifts towards integration of AI-enabling talent into industries through the change in the share of AI-enabling jobs within an industry cluster over the 2010-to-2018 period. If industry clusters are incorporating more AI-enabling capacities into their business models, one would expect to observe significant increases in the proportion of related employees active in those industry clusters over time. For Indiana, there does not yet appear to be overwhelming evidence of these types of shifts occurring in recent years as most industry sectors with significant AI-enabling occupational employment levels above 5 percent of the total industry employment collectively average only a 1.7 percentage point increase in the share of these occupations. Several industry clusters with significant employment footprints do appear to be building towards more significant shifts in occupational mix, in particular advertising, banking, and credit-related financial services clusters which average approximately a four-percentage point shift in occupational mix towards AI-enabling occupations.

Figure 14: Presence of AI-Enabling Occupations in Leading Traded Indiana Industry Clusters, 2010-2018



Note: occupations with >7,000 Indiana employees in 2018 annotated.
Source: EMSI Inverse Staffing Patterns, EMSI 2019.4

Regional education pipelines are an important resource that Indiana employers can leverage if they want to shift their occupational mix towards AI-related talent in more significant ways. The state's research universities are generating significant volumes of graduates that are aligned with the types of occupations required to advance AI-enabled applications for industry. Understanding the balance between the state's industrial labor force demand and the supply of skilled talent from universities can provide an indication of the extent to which the talent pipeline may be able to meet the needs of regional companies and provide capacity for growth.

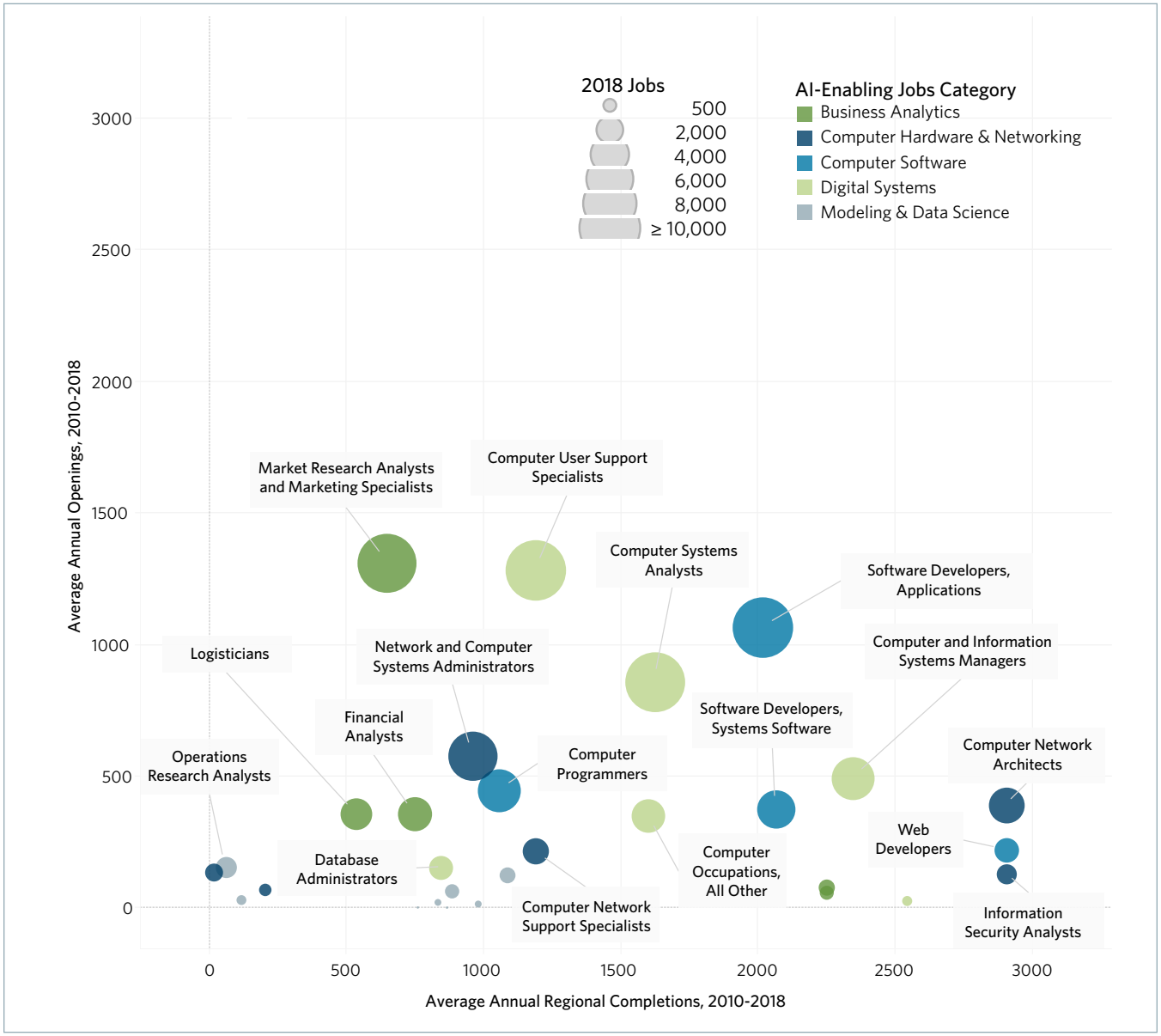
Utilizing IPEDS and EMSI data on openings and completions that crosswalk degree types to occupational categories that require highly similar technical skills and knowledge, Figure 15 compares the "supply" of degree completions from regional institutions to the "demand" from industry openings over the 2010 to 2018 period in the AI-enabling occupation categories. While this data does not account for particular skills matching or quality of graduates, it can

provide some insights into potential shortages or surpluses of talent volume. As an illustrative example, consider the detailed occupation of computer systems analysts which is a part of the broader digital systems AI-enabling jobs category. In 2018, Indiana employment in this occupation stood at 9,787 total jobs. Average annual completions in degree programs over the 2010 to 2018 period for the region totaled 1,624, while average annual job openings within Indiana industries over the same period totaled 859. These levels indicate that there was a "surplus" in the "supply" of individuals completing degrees aligned with this occupation relative to the "demand" from industry as signaled by the average number of annual openings.

For most occupations in AI-enabling categories, Indiana's average annual degree completions volume from 2010-to-2018 exceeds the number of average annual openings in occupations for which those degrees are highly aligned. One notable exception in occupations that have high employment footprints in the state includes market research analysts. For this occupation, the number of openings from 2010-to-2018 were significantly higher than the number of degree completions, indicating a potential shortage in talent generation from higher education institutions. While this indicates that Indiana's higher education institutions have the capacity to train a supply of talent that can satisfy industry openings in AI-enabling occupations, it would require a level of retention of these graduates in excess of what is currently occurring.

Across many other segments of AI-enabling jobs, there is evidence of a surplus of degree completions that outpaces openings, indicating a robust pipeline available for industry to leverage. However, feedback from industry stakeholders makes it clear that a lack of skilled talent is a key issue that companies are trying to address in order to build capacity. This difference in narratives indicates that despite a significant talent pipeline, there is still a mismatch in the labor market for AI-enabling occupations within the state being driven by factors outside of capacity. There are several potential explanations which could point to issues of a skills mismatch in degree programs to state industry needs, a recruitment and occupational mix mismatch on the part of companies in seeking the right kinds of skilled talent, or labor market dynamics that lead regional graduates to pursue openings outside of the state. These trends could also be driven by the need for experienced rather than entry-level talent to fill industry roles, with companies unable to meet demand because recent graduates have not yet had significant exposure to business operations. In addition, IU, Purdue and Notre Dame attract a significant number of students nationally and internationally as each are recognized as top institutions for training in these related areas. Conversely, the companies in Indiana that are driving demand for these graduates, while national and global leaders in their industries, are not recognized as drivers of innovation in AI and are therefore not as attractive as brand names such as Amazon, Google, or even expert consulting firms such as McKinsey, Deloitte and the like. Two key occupations that have significant employment footprints within the state as well as indications of a surplus of talent relative to openings are computer systems analysts and software applications developers. These two job types are key enablers of the IT infrastructure and applications development processes required to enable a robust AI technologies ecosystem and are potential targets for further investigation to identify the underlying conditions driving the mismatch in supply and demand.

Figure 15: Supply-Demand Balance for AI-Enabling Occupations in Indiana, 2010-2018



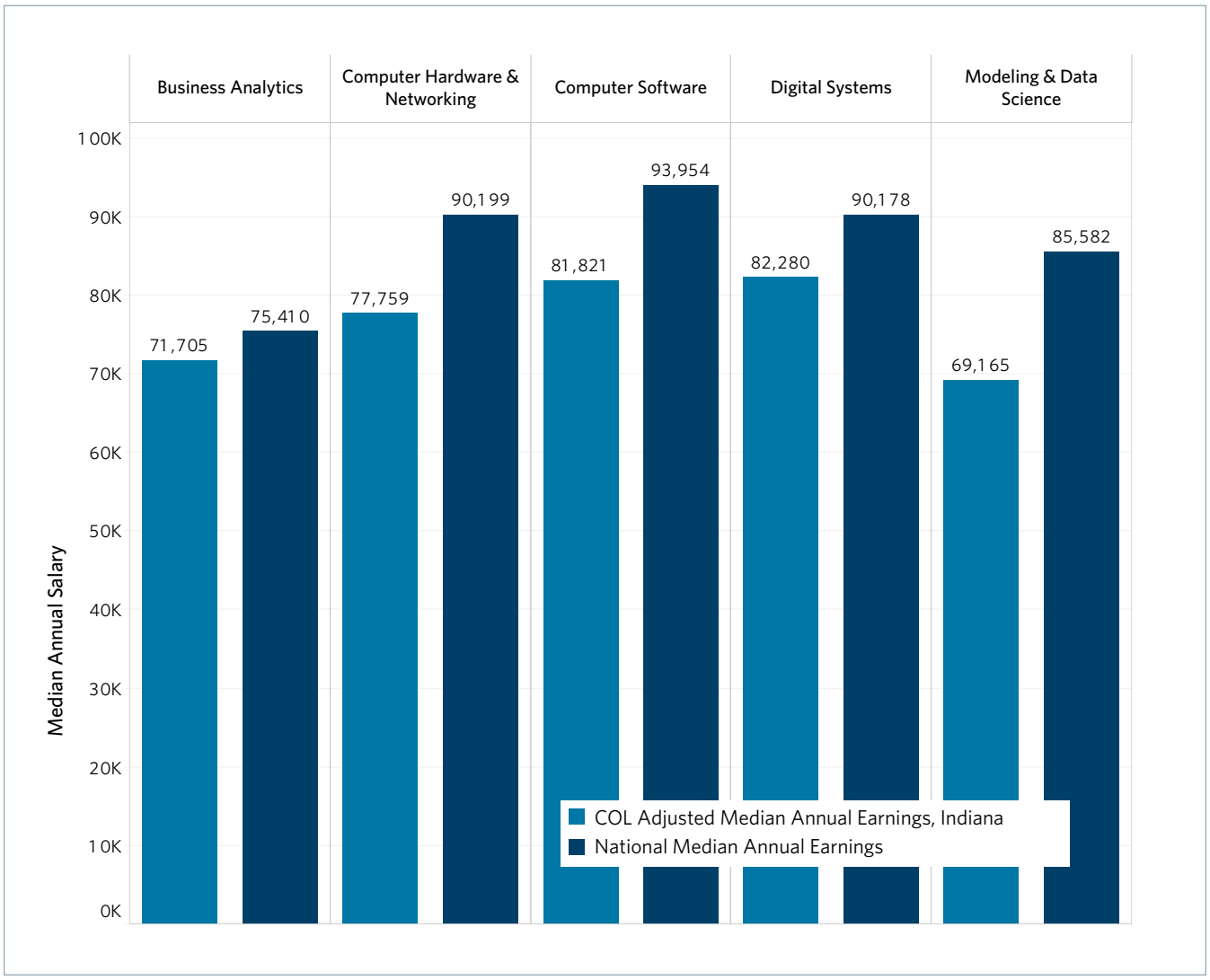
Note: occupations with >1,000 Indiana employees in 2018 annotated
Source: EMSI Occupational Data, EMSI 2019.4; NCES IPEDS Data

One important factor that determines the employment decisions of skilled talent is competitive salary levels. For the AI-enabling occupational space, this aspect of attracting talent is particularly important because of a highly competitive national labor market for these skill sets as well as increased salary expectations that are set by prominent employers in leading technology innovation ecosystems in Silicon Valley and other major coastal metro areas.

Figure 16 shows the national median annual earnings levels across major AI-enabling occupational segments as well as the cost of living adjusted salaries for Indiana. The overall national median annual earnings level of $84,772 for AI-enabling occupational segments is 21.4 percent higher than the cost of living adjusted Indiana median annual earnings level of $69,850, meaning that **while labor cost for talent is lower in Indiana, the difference in actual purchasing power relative to national trends potentially makes the state a less attractive destination for employment in these spaces.** National salaries are consistently higher than cost of living adjusted salaries for Indiana across key segments

and occupations, potentially providing some insight into why industry needs are not being met despite signals of talent pipeline surpluses.

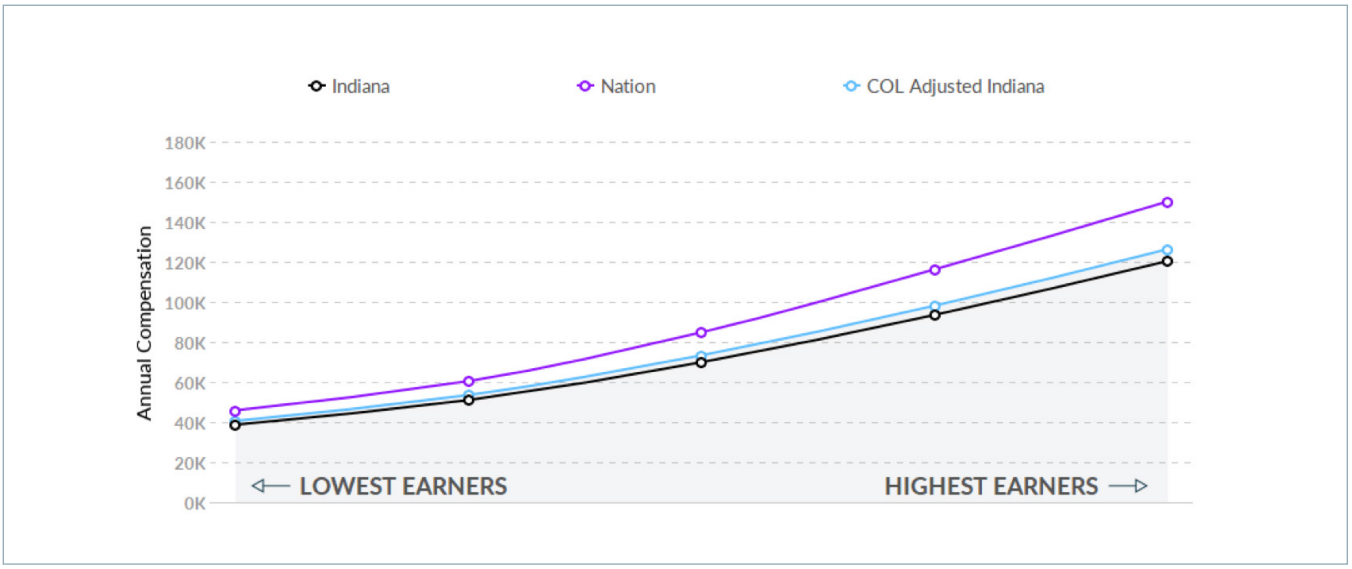
Figure 16: Median Annual Salaries for AI-Enabling Occupational Segments in Indiana, 2018



Source: EMSI Occupational Earnings Data, EMSI 2019.4

Figure 17 provides additional evidence that salary levels for AI-enabling occupations offer lower market value compared to national evidence. The annual earnings distribution shows that cost of living adjusted median annual earnings in Indiana are consistently lower than overall U.S. levels even at lower earning percentiles and that lower labor costs are not offset by cost of living advantages. This presents a significant challenge for Indiana companies seeking to retain portions of the talent pipeline as well as attract top talent from other regions of the country.

Figure 17: Distribution of Median Annual Salaries for AI-Enabling Occupations in Indiana and U.S., 2018



Source: EMSI Occupational Earnings Data, EMSI 2019.4

Job postings offer a more refined lens and specificity that complements the preceding analysis utilizing federal occupational classifications that are often more general in nature. EMSI maintains a job postings tool that allows for examining actual job titles used by companies, and summarizes the skills requirements described in the postings. Using job postings data collated by EMSI from leading job postings websites from September 2016 to April 2019, it is possible to get additional indications of the leading areas of demand for skilled talent in the state. Table 19 shows the top job titles in job postings for the state over this time period.

The highest frequency of job postings is focused in computer and IT positions, with a particular emphasis on software engineering and back end information technology support. Applied analysts and managers are also significant areas of activity in position postings.

Table 19: Top Job Titles in Job Postings Data for Indiana Active Postings, 2016-2019

Job Title	Unique Postings from Sep 2016 - Apr 2019
Software Engineers	10,171
Business Analysts (Business and Financial Operations)	5,780
Java Developers	4,368
Project Managers	3,081
Research Analysts (Life, Physical, and Social Science)	2,969
Financial Analysts	2,949
Business Analysts (Computer and Mathematical)	2,844
IT Help Desk Specialists	2,189
.Net Developers	1,999
Systems Analysts	1,969
Systems Engineers	1,933
Systems Administrators	1,762
IT Support Analysts	1,650
IT Quality Assurance Analysts	1,567
Web Developers	1,543

Source: EMSI Job Posting Analytics, EMSI 2019.4

Similarly, the top hard skills listing in Indiana job postings for AI-enabling occupations are listed below in Table 20. Skills related to software development are again very prevalent, but also noteworthy is the frequency of business requirements and business process skills which indicate an applied context for many positions.

Table 20: Top Hard Skills in Job Postings Data for Indiana Active Postings, 2016-2019

Skill	Frequency in Postings (167,685 total postings)
Microsoft Access	19%
SQL (Programming Language)	17%
Agile Software Development	13%
Java (Programming Language)	11%
Automation	9%
Software Development	9%
Business Requirements	9%
JavaScript (Programming Language)	9%
Accounting	9%
Project Management	9%
Auditing	9%
Software Engineering	9%
Business Process	8%
Operating Systems	7%
Information Systems	6%
Linux	6%
Technical Support	6%
Cascading Style Sheets (CSS)	6%
.NET Framework	6%
C# (Programming Language)	6%
Python (Programming Language)	5%
Scripting	5%
Scrum (Software Development)	5%
Systems Development Life Cycle	5%
HyperText Markup Language (HTML)	5%
New Product Development	5%
Web Services	5%
Application Programming Interface (API)	5%
Customer Relationship Management	5%
Debugging	5%

Source: EMSI Job Posting Analytics, EMSI 2019.4

D. Entrepreneurship and Growth of an Advanced Analytics Service Sector for Indiana

In addition to university R&D core competencies and capabilities and more mature industry-leading companies, a robust and high-functioning advanced analytics ecosystem is complemented by innovative startups. These new and emerging firms play a vital role with respect to product and service development, as well as providing solutions, not only locally but often serving a national or even global marketplace. To more fully inform Indiana's asset base in advanced analytics capabilities from this industrial vantage, as well as to identify where there are concentrations of emerging firms and competencies but also significant gaps relative to other states, both a State of Indiana and national analysis of venture capital (VC) investments were conducted.

Investment by VC firms in emerging companies is just one of many potential metrics that can be used to gauge innovation and technology trends; however, venture investment represents a significant milestone in the commercialization process where capital is being actively deployed to fund innovative concepts that have high growth potential. Using investment in emerging, innovative companies as an indicator of innovation is especially relevant for AI and advanced data analytics technologies for the following reasons:

- Companies receiving VC investments will tend to be more focused on the market applications of AI and advanced data analytics that give insights into practical usage rather than "basic research";
- VC investment has been concentrated in tech industries in recent time periods, making it a critical resource for commercializing emerging AI-related applications; and
- AI and advanced data analytics technologies are often not widely documented in other innovation indicators (e.g. patents/IP) due to a desire to maintain a competitive advantage in a crowded private market.

Overview of U.S. VC Investment in AI-related Companies and Indiana's Position

Using the PitchBook database, the analysis collected detailed information for all U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding from 2014 through 2018. A total of 1,671 companies were identified in AI/machine learning applications verticals who have raised a total of $71.8 billion over the last five years.

As shown in Table 21, most AI-related companies attracting significant funding are oriented towards business and enterprise software products, as one would expect in the AI space. The notable exception in terms of industry applications is the transportation industry, which reflects the success of Uber in generating extremely large investment in its pre-IPO stage ($19.9B for Uber, not shown in the analysis, to avoid its distortion effect).

Table 21: VC investments in U.S. AI-related Companies by Industry/Application Area, 2014-2018

Primary Industry Sector/Application Area	Number of AI-related Companies	Total Raised to Date ($M)
Business/Productivity Software	365	$7,046
Application Software	192	$2,985
Media and Information Services (B2B)	85	$1,735
Network Management Software	81	$3,654
Automation/Workflow Software	76	$2,570
Social/Platform Software	74	$1,695
Other Software	66	$1,324
Database Software	65	$5,736
Financial Software	57	$1,692
Other Healthcare Technology Systems	48	$1,092

Source: TEConomy's analysis of PitchBook Venture Capital Database.
Note: the analysis is limited to U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding during the 5-year period. The large investment in Uber of $19.9 billion is not included.

On a state level, California dominates AI-related VC investment activity, followed by east coast tech hubs in New York and Massachusetts (Table 22). Indiana ranks near the middle among all states in the count of companies and total funding; but the state is well below leading states in AI-related private equity investment.

Table 22: VC investments in U.S. AI-related Companies for Leading States and Indiana, 2014-2018

State	Number of AI-related Companies	Total Raised to Date ($M)
California	856	$54,163
New York	194	$6,247
Massachusetts	140	$3,619
Texas	75	$1,032
Washington	62	$893
Virginia	33	$712
Illinois	32	$1,912
Pennsylvania	32	$505
Florida	29	$198
Indiana	**5**	**$26**

Source: TEConomy's analysis of PitchBook Venture Capital Database.
Note: the analysis is limited to U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding during the five-year period.

Indiana is home to five AI-related companies that received a significant ($0.5 million or greater) VC investment during the five-year period (Table 23). The companies, which combined to raise $26.1 million to date, largely reflect the established industry and tech strength of this space in Greater Indianapolis around cloud-based digital marketing.

This reveals the limited context in which Indiana AI-related companies are currently operating; and while an important strength that centers around the key role that Salesforce plays in the regional and state economy, this market niche offers relatively little in applicable solutions (beyond digital marketing) for corporate solutions and broader business functions demanded by companies in Indiana's leading clusters in advanced manufacturing and health and life sciences.

Table 23: VC investments in Indiana-based AI-related Companies, 2014-2018

Company Name	Company Description	Primary Industry Sector/ Application Area	Total Raised to Date ($M)
Pattern89	Developer of a cloud-based marketing platform designed to help digital marketers to experiment at scale	Automation/Workflow Software	$7.50
DemandJump	Provider of a cloud-based marketing intelligence platform intended to make marketing more purposeful.	Media and Information Services (B2B)	$8.55
PERQ	Provider of artificial intelligence technology for consumer and website engagement and conversion.	Business/Productivity Software	$7.23
OpenMartech	Developer of a SaaS software designed to assist in analyzing marketing sales and customer insights using artificial intelligence.	Business/Productivity Software	$1.95
FWDNXT	Developer of a mobile co-processor designed for accelerating deep neural networks effective at image recognition and classification.	Electronic Components	$0.90

Source: TEConomy's analysis of PitchBook Venture Capital Database.
Note: the analysis is limited to U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding during the five-year period.

Jobvite – A Successful Example of Private Equity Funded Analytics Growth Company in Indianapolis

Jobvite, an employer-employee search engine company, began in Indianapolis under the name Canvas (developing chat-based tools, for HR applications, filtering recruitment candidates). Canvas was acquired by a California-based private equity firm, that rolled up Canvas with three other acquisitions and then located/relocated the resulting restructured merged company Jobvite in Indianapolis, where it is now headquartered and growing (at close to 100 employees in the Fall of 2019).

Using Network Analysis to Understand the Landscape of Emerging AI Technologies

The national VC investment landscape offers the opportunity for a network analysis to better understand the full landscape of emerging AI technologies. The analysis utilizes machine learning and natural language processing to understand the contextual nature of AI companies beyond their basic industry classifications, and to identify common technology and applications area "themes" seen nationally. Networks help to visualize the relationships between the technologies and market applications areas of companies based on the unstructured text content of descriptions of companies receiving VC investments. Comparing similarities in this text content yields "connections" between companies which build out the network of relationships within the AI space. Community detection algorithms can

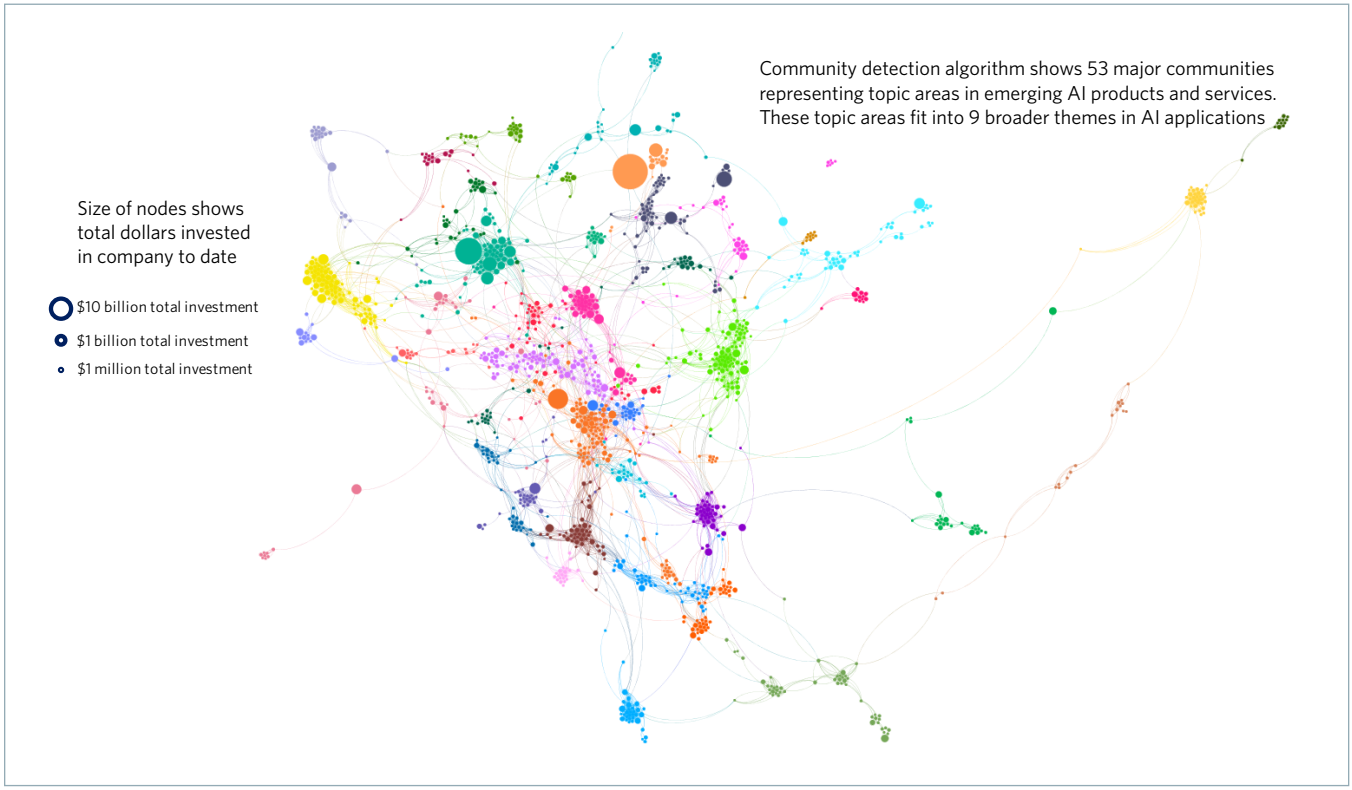
use the structure of this network to identify clusters of highly connected companies that have common technology or market application themes.

The algorithms identified 53 major "communities" representing a breadth of topic areas in emerging AI-related products and services nationally, fitting into nine broader thematic areas in AI applications, including:

- Computing and AI Tools,
- Autonomous Systems,
- Business Services Analytics & Process Automation,
- Marketing & Content,
- Education; Entertainment & Consumer Goods/Services,
- Biomedical,
- Imaging,
- Infrastructure, Industrial Controls, & Environment, and
- Security.

Figure 18 shows the high-level depiction of the network analysis. For individual thematic depictions, see the additional graphics in the Appendix to this report.

Figure 18: Innovation Landscape Network of U.S. AI Companies Receiving Significant VC Investment, 2014-2018



Source: TEConomy's network analysis of PitchBook Venture Capital Database firms.
Note: the analysis is limited to U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding during the 5-year period.

Indiana's position in this national AI-related innovation ecosystem is extremely limited, with a niche strength identified in the broader Marketing and Content thematic area identified nationally. Its larger industry cluster context aligns more closely with other innovation activity in the AI space underway nationally, such as biomedical applications, industrial process automation, and security applications, to name some. This provides one direct indication as to why Indiana firms are often looking elsewhere for corporate-led AI-related innovation and solutions. Going beyond the venture-backed AI-focused companies, however, does reveal a set of emerging companies in related advanced analytics areas, some of which are showing great promise.

Broader Look at Emerging Indiana Companies in Advanced Analytics Advising and Services

Beyond the more restricted lens for AI-focused analysis of emerging, venture-backed companies, there are indeed additional emerging tech companies across Indiana providing customized consulting solutions to customers needing advanced analytics advice and services. These companies include:

- **Aunalytics** in South Bend. Contract AI/analytics services and consulting focused on mid-sized companies. Aunalytics is able to integrate tailored IT solutions, cloud infrastructure along with Big Data, analytics and AI solutions. A high-growth company that is attracting Indiana and external AI and advanced analytics talent into South Bend.
- **Data2Discovery, Inc.** in Bloomington. The firm offers a platform that includes linked data, graph technologies, machine learning and AI, working to solve business challenges in drug discovery.
- **Digital Health Solutions** in Indianapolis. Has developed integrated clinical decision support technology to engage patients, families, and clinicians to improve health care delivery with a focus on children.
- **hc1** in Indianapolis. Utilizes clinical lab data to develop actionable health care insights, enabling personalized care for health systems and diagnostic labs.
- **Indigo BioAutomation** in Carmel. The firm is using software automation and human-computer teams to transform lab operations, particularly with respect to fully automating chromatography and mass spectrometry quantitative analysis.
- **LifeOmic** in Indianapolis. Utilizing innovative cloud and mobile software to advance precision health, with focuses in oncology and cardiology solutions.
- **Megaputer** in Bloomington. Provides key insights using data mining and text analytics to inform better decision-making utilizing expertise in linguistics and machine learning approaches.
- **MedeoLinx** in Indiananapolis. Develops informatics solutions and analytics in the life sciences including in drug discovery and diagnosis. The firm combines strengths in next generation sequencing, data mining, and network pharmacology for potential drug targets in addition to biomarker research.
- **Springbuk** in Indianapolis. Applies a health intelligence platform to transform health data into valuable analytics and actionable recommendations to strategically inform and manage employee health benefits.

Exact Target Acquisition by Salesforce – Indianapolis-Based Entrepreneurial Enterprise Results in Substantial Growth in the Indianapolis Presence of Major Digital Enterprise and Wealth Creation Event Sparking Further Entrepreneurship and Investments

ExactTarget was founded in Indiana in late 2000 by three entrepreneurs, focusing on digital marketing automation and analytics software and services. The firm experienced dramatic growth and executed a successful initial public offering in 2012, by which time the firm employed 1,500 personnel. The success of ExactTarget attracted an Acquisition of the firm by Salesforce for $2.5 billion in 2013 – representing the largest acquisition, by value, of a tech firm in Indiana's history.

Salesforce has continued and grown its operations in Indiana and is a key employer in Indianapolis. The large liquidity event for investors in ExactTarget, and the entrepreneurial and business growth experience of the firm's leadership, has been the impetus for ongoing growth in regional business start-ups and early stage risk capital.

E. Supporting Organizations, Programs and Events

If we view advanced analytics, data sciences and AI as elements in a digital technology ecosystem, then it is evident that Indiana has established and is further growing a system of ecosystem supports.

The Central Indiana Corporate Partnership (CICP) is an example, working on coordination and connectivity activities for technology-based economic development. CICP and its individual initiatives are designed to provide cross-cutting ecosystem support services and programs in areas such as talent access, and has focused programs directed towards key advanced industry cluster verticals.

Advanced analytics, data sciences and AI are relevant to the current and future development of each of Indiana's advanced industry clusters, and each of the individual CICP initiatives is recognizing the need to engage in advancing these capabilities for the region and the state:

- **AgriNovus Indiana** – where agbioscience and agricultural technology developments are engaging advanced analytics and AI in diverse areas such as genetic and genomic analysis, precision and digital agriculture technologies, and business analytics.
- **Ascend Indiana** – CICP's talent focused initiative, working to build a network whereby employers and talent can find and connect with one another.
- **BioCrossroads** – where stakeholders ranging from biopharmaceutical companies to health systems to Indiana's research universities are deploying advanced analytics for activities in drug discovery, clinical research, healthcare informatics and health system management.
- **Conexus Indiana** – which is working to enhance and support advanced manufacturing and logistics industries in the region, industries that are at the heart of automation and AI-driven advancements.
- **Energy Systems Network (ESN)** – where AI and other advanced data sciences applications impact opportunities in areas such as autonomous vehicles, SMART infrastructure management and SMART grid implementation.
- **TechPoint** – CICP's IT and digital ecosystem organization, working to assure that the IT and digital tech sector in Indiana has access to talent, R&D-based innovations, capital, and associated entrepreneurship and business development and growth services.

In the 2018 "Clusters and Disruptors" report for Central Indiana, TEConomy concluded:

It is clear that the general structure and individual initiatives of CICP are well-suited to the current cluster profile of Central Indiana. Each of the six clusters identified via TEConomy's analysis finds a natural fit within the current mission and focus of at least one of the CICP initiatives. This "cluster vertical" orientation of CICP initiatives has served the region well. Where change may be necessary at CICP relates to the discussion of convergence and the impact of disruptive technologies and modern business megatrends that are impacting, and will increasingly do so, the Central Indiana economy.

Digital technology and associated data analytics are at the heart of most convergence opportunities and it will be particularly important for TechPoint to examine ways to leverage Central Indiana's fast-growing IT sector to encourage connectivity to the verticals under AgriNovus, BioCrossroads, Conexus and ESN. IT is not the only convergence hub, however, and cross-cluster opportunities are likely to exist in multiple other arenas, for example:

- *Experience with early adoption of robotics in the advanced manufacturing (especially automotive) and the logistics cluster feed into opportunities in robotics in agriculture.*
- *Biosensor applications in agbioscience translate to other life sciences applications, and vice-versa.*
- *The application of AI advancements and associated expertise, in whichever cluster they appear first, are germane to almost every other cluster.*
- *Convergence across life sciences and agbiosciences in terms of molecular and genomic sciences, big data analytics, advanced diagnostics, biopharmaceuticals and devices. The IBRI is likely to be a key institution in considering convergence opportunities here and making connections (supported by CICP).*

CICP already has notable experience in facilitating convergence through programs developed by ESN. Indeed, ESN was established to combine innovation and collaboration across energy technology and application arenas – serving as a convergence hub to bring together multiple stakeholders to address energy systems opportunities in areas such as mobility, energy efficiency, energy storage and plug-in hybrid solutions. As noted by Paul Mitchell, ESN President and CEO:

> *"No one company has the answer. Solutions require collaboration and a systems approach.[64]"*

ESN is, in many respects, a model for what CICP needs to achieve in terms of preparing for and facilitating cross-cutting initiatives to address convergence opportunities. ESN has a special focus on facilitating the formation of multi-company collaborations designed to accelerate new technology commercialization through commercial scale pilot projects. This focus on moving collaborative projects off of the "bench" and into pilot scale testing is particularly noteworthy – ensuring that new collaborations are very much focused on scaling innovations with an eye to rapid commercialization. ESN (and other CICP initiatives) are also experienced in convening workshops and events around new technologies, and this will be an important function to continue and expand as the industry focus groups expressed substantial interest in events that would highlight cluster challenges, needs, and opportunities that may benefit from cross-cluster collaborations.

Overall, TEConomy concludes that CICP is well structured to continue to advance technology-based economic development in Indiana. Moving forward, however, more attention will need to be paid to standing up formal structures and programs to facilitate cross-cluster convergence and be strategic in approaches to megatrends and disruptive technologies.

64 http://energysystemsnetwork.com/

The conclusions from 2018 are only reinforced by the current advanced analytics project. **It should be noted that there is even more urgency for developing the right model for accessing cross-cutting data sciences, advanced analytics and AI talent.** Data science, advanced analytics and AI are central to advancing Indiana's clusters, and broader industry fortunes, that collaborative approaches are necessary to:

- Integrate CICP's current work in connecting talent to cluster industries from across CICP's various initiatives, and
- Provide a roadmap for making connections between industries and their needs and faculty and research and consulting services within Indiana's research universities.

It is important to note that CICP's initiatives have already structured multiple programs that are enhancing talent-to-employer and employer-to-talent connectivity in Indiana. These existing programs are evident in both the programming of sector specific initiatives (such as AgriNovus in agbiosciences and TechPoint in IT domains) and cross-cutting through the dedicated talent access programs of Ascend Indiana. These existing programs are well-suited to being tied-in to support enhancement of talent access in the advanced analytics and AI space. Major Indiana employment sectors with significant demand for advanced analytics talent will most likely already be engaged with one or more of the CICP initiatives as are the major universities in the state. Examples of programs well-structured for addressing data analytics connectivity for talent and employer needs include:

- **The Ascend Network** – Operated by Ascend Indiana, the Ascend Network is a software enabled system that operates with front ends accessed by potential employees and by potential employers. There are currently 280 employers on the network with profiles for the job characteristics for 745 unique internships and job categories. The Network is successful because representatives of Ascend meet with potential job seekers and help develop accurate profiles of their skills, education, work characteristics and what they are looking for in a job and working environment. To-date, upwards of 5,000 candidates have been incorporated into the system, and employers are finding the service to be very well aligned in matching candidates to their needs.

- **TechPoint** – The Xtern program operated by TechPoint was launched seven years ago and focuses on connecting talented college students to Indiana technology focused companies with a need for their training and skills. Xtern rigorously vets applicants for its intern placement program, assuring that employers gain access to well-qualified developing talent. The program receives applications from students nationwide (in 2019 this comprised of 1,900 applications from students at 212 universities and 45 home states) and for 2019 accepted 209 students for the Xtern Finalists Day, where the Indianapolis Convention Center is used to host individual meetings between the applicants and over 60 companies. Xtern represents an innovative and proactive approach to building awareness for employment opportunities in the Indianapolis area tech industry and building a positive image for the state with tech-savvy students from across the nation.

- **AgriNovus** – "Field Atlas", to be launched in early 2020, is a developing AgriNovus program that focuses on career discovery and exploration. Designed as a Gen Z tool, Field Atlas targets both high school and college level students, providing an interactive digital information tool that profiles cutting edge and emerging jobs in agricultural bioscience and ag technology sectors. The program will allow students to explore more than 80 ag sector careers, seeing why the job matters and what typical days of work and work challenges entail. It allows exploration of relevant college majors and program of study offered at Indiana higher education institutions. AgriNovus, in collaboration with Ascend Indiana, is also undertaking a skills gap assessment and is in direct discussion with employers to discern skills and education areas that companies are finding to be lacking.

In addition to, and complementing the work of CICP initiatives, several organizations provide connection services, particularly through events and associated networking opportunities. Examples include:

- **Indiana University Network Science Institute (IUNI) Events:** IUNI hosted the Conference on Big Data and Network Science in 2017, and it organizes regular NetSci Talks and Open Science Forums. All NetSci talks are live-casted and archived.
- **Indy Big Data:** The annual conference attracts 1,000+ attendees to workshops and talks on everything from analytics and data management to security and IoT.
- **IUPUI School of Informatics and Computing Colloquia Series:** IUPUI invites a variety of leaders in informatics research and practice to campus. Talks cover a range of topics, but healthcare is often emphasized.
- **Kelley School of Business Institute for Business Analytics (IBA) Conferences:** The IBA has multiple conferences at IU, including forums devoted to specific aspects of analytics (e.g. marketing, supply chain, finance, etc.).
- **The Purdue Integrated Data Science Institute (IDSI) Summit** was held in November 2019 and will be an annual event providing the university community and industry with a full day of programming and national speakers around data sciences topics and joint opportunities.
- **Purdue IDSI** maintains an active program of events, regularly updated on their Website, focused on data sciences topics. Examples include a seminar series on data sciences and digital agriculture and an open seminar series in critical data studies.

TechPoint currently hosts and maintains an events and "signature events" calendar on its website and it may be beneficial to have a page dedicated to specific analytics-oriented events, workshops and conferences.

IV. Strategic Recommendations

A. Headwinds Facing Indiana in Building an AI-Enabling Ecosystem

Indiana faces a number of emerging trends as it seeks to continue to expand its research and talent activities, including broader market trends as well as challenges that are unique to the state's ecosystem. Alongside the state's university assets and industry needs described in the previous sections, these background trends help set the context for identifying the right fit for a model of public-private engagement for the state as well as strategic priorities for investment and intervention.

Indiana's major corporations operate on a global scale and their investments, including investment in advanced analytics capabilities, are taking place in multiple national and international locations. Advanced data analytics and AI have been, and are, such fast moving spaces that Indiana companies have evidently sought access to talent, consulting services and partnerships at hubs outside of Indiana.

Yet, as shown in this report, Indiana has significant talent development and research focused institutions in advanced analytics and AI. The current challenge is to urgently connect the evident capabilities and capacities (that largely reside within Indiana's major research universities) with in-state industry to lessen the flow of investments being made outside of the state.

Indiana's institutions and companies are in the midst of transitioning from legacy models of research and business operations towards digital, analytics-powered research and applications. This change is occurring against the broader backdrop of similar national and global transitions as major markets seek to identify and integrate advanced analytics solutions in a rapidly-moving environment. The challenge for Indiana has been that major enterprises in the state have been tending to look outwards for solutions to their analytics partnership and talent access needs. Many of these companies already have major operations in other national and international locations, including presence within established and fast growing AI and advanced analytics geographic hubs such as New York, California and Massachusetts. Major companies have choices as to where they invest and chose to grow, and if Indiana and the Central Indiana region are not responsive to building an AI and advanced analytics ecosystem suited to enterprise needs, the risk is very real that their attention will be focused elsewhere. While such external investment would represent a potential lost opportunity for the growth of AI and advanced analytics capacity in Indiana, it is a less than ideal for another reason. As Indiana companies work to establish a presence in out-of-state AI and advanced analytics hubs, they will be operating in competition with large numbers of other enterprises seeking to access the same talent pool within those hubs. It is thus apparent that investing in building a more robust, complete and scalable AI and advanced analytics ecosystem in Indiana is associated with multiple potential economic benefits:

- It will help reduce outward investment by Indiana enterprises and retain their investment dollars in-state,
- It will help Indiana enterprise better access talent by reducing a dependency on competing for highly sought-after talent in the existing hubs, and

- It will provide opportunities for Indiana to better retain its AI and advanced analytics university graduates and connect them with Indiana employment opportunities.

Many of the challenges facing Indiana are also shared by broader national research and industry ecosystems and their stakeholders as they navigate a period of disruptive change, making it important to distinguish these larger scale trends from more regional and local challenges. Several high level market headwinds in advanced analytics and AI-enabling technology areas that are important for the state to adapt to in its efforts include:

- The integration of data sciences and the infrastructure that enables collection of large data streams are becoming necessities for modern businesses, regardless of industry sector;
- Accessing and incorporating innovations in AI-related technologies requires a fundamentally different paradigm than traditional commercialization models – innovation in this area is extremely fast-paced and utilizes virtual, distributed, and open source assets;
- As the enabling technologies continue to become cheaper and more accessible, pipelines of skilled talent able to utilize the tools of data sciences to solve business problems are the key to competitive advantage; and
- Issues related to ethics, privacy, and bias linked to the deployment of technologies in this space are becoming increasingly critical for businesses to consider.

Many of Indiana's leading research institutions are beginning to pursue high quality research and capacity-building efforts that seek to address these larger trends, and while the state's industry leaders are at various levels of progress in their transition towards analytics-driven business models they collectively recognize the importance of not falling behind the larger market direction and the role of talent as a key enabler of success.

In addition to the broader market headwinds driving the direction of efforts to grow AI-enabling capabilities, Indiana also faces several key challenges as a region in a competitive national and global market. These challenges include factors such as:

- Established coastal AI ecosystems already exist with significant critical mass and resources, making competition in fundamental "basic science" AI research and development difficult;
- Ecosystems that produce significant volumes of high-quality talent will have the biggest competitive advantage in attracting additional critical mass to their environments, and Indiana's talent pipelines are currently not well connected to in-state industry; and
- Indiana's industry stakeholders represent a wide array of different industry sectors, technology readiness levels, and desired scopes for engagement when considering deployment of AI-related technologies. However, a successful regional ecosystem must be able to align these three dimensions around specific solutions in order to achieve critical mass and drive investment.

Through purposeful design of strategic initiatives, Indiana has an opportunity to directly address these regional challenges and improve the state's competitive position relative to peer regions who are advancing their own analytics ecosystems.

Indiana's current position in AI-enabling research and talent generation tends to reflect the difficulty in meeting these challenges without a coordinated effort across institutions and industry stakeholders to scale a community of practice in advanced analytics. **Currently, most of Indiana's skilled graduates flow to out-of-state employers, and out-of-state companies are aggressively engaging talent pipelines early in their education and embedding their presence on university campuses.** This reflects the reality that Indiana's data sciences graduates are in high demand in a competitive national labor market and presents a particularly important challenge to address that is one of the drivers of current misalignment in talent pipelines.

Indiana industry stakeholders have also largely focused their efforts to meet their needs for advanced analytics services and talent outside of the region, although some very recent shifts towards significant engagement with in-state university partners has started to occur. Most university engagement with Indiana industry today is at smaller scales and revolves around direct relationships with key centers or faculty rather than broader university capabilities or initiatives, with major university corporate partners also tending to be out-of-state. In addition, **many university initiatives seeking to grow AI-enabling talent flows and capabilities are relatively new or only recently being brought up to scale, meaning that awareness on the part of regional industry stakeholders is also an issue to be addressed.**

The overarching narrative outlined by the convergence of these factors as well as the state's current industry and research position all highlight a missing element for Indiana in growing the regional footprint and competitive position in AI-related spaces. **This need can best be described as the development of a shared community of interests involving corporate and university partners, through intentional coordination and communication initiatives, that realigns talent flows and pushes in-state industry needs to universities in a broad rather than narrow way.**

B. National Models for Structuring Public-Private Engagement Around Data Sciences

A number of different potential models for engaging public-private collaboration around data sciences and AI-related technologies have emerged across the country that can help serve as models for Indiana's strategic path forward in growing a community of practice in AI-related applications.

No single model of structuring engagement is dominant as a best practice, and each type of model has grown out of a desire to adapt to the needs of existing industry partner demand in a very tailored way. The various types of models can be summarized around their membership and organizational structures as follows:

- **Industry Affiliates Portal/Single Access Point:** Utilizes a dedicated university function to coordinate industry partnership and engagement activity through a single point of access to broader university capabilities in data sciences, either on a case by case basis or via a university affiliates program. The university office or program typically takes responsibility for matchmaking industry needs to more specific faculty or departments within the university.

- **Embedded Applied Research Company/Proximity** – Leverages an applied research company or consultancy that is housed directly on a university campus towards engaging industry partners through contracts for services. The company is typically a separate business institution, but often leverages the proximate university faculty as senior advisors with a staff of dedicated applied researchers that carries out specific industry projects.

- **Industry Membership/Advisory Board** – Uses a membership model where industry partners pay for preferred access to university data sciences capabilities, research, and talent. The industry membership or advisory organization typically connects through an interdisciplinary university institute or center that can leverage additional university capacities as necessary, and industry partners typically get benefits such as curriculum advisory roles, early access to new research, access to up-skilling opportunities for their workforce, or priority recruitment opportunities for students.

- **Centralized Innovation Hub** – Establishes a central physical facility that aggregates industry interaction with emerging university research and university engagement with applied industry needs in a single location. The hub is typically grounded around a mixed-use development that includes space for industry partners as well as research assets (such as labs, IT infrastructure, etc.) and is at least partially sponsored by industry partner investment.

- **Data sciences Environment (DSE) Industry Pull** – Establishes a dedicated environment for talent generation on university campuses that incorporate immersive educational and technical skills programs in data sciences, industry interaction and exposure to business models, and capstone projects that drive development of talent attuned to specific industry partner needs. Industry partners typically sponsor cohorts of students progressing through the program and engage them around applied needs throughout their degree programs, while students come from a variety of discipline areas and participate in a competitive process for "residency" to the DSE program.

- **Outward Push/Grand Challenge** – Uses high level problem statements targeted towards specific scientific and technical innovation goals to solicit participation by collaborative teams in leveraging grant funding towards development of commercialization outcomes in data sciences applications. Problem statements usually encompass high level thematic areas, and are set by the funding organization (in most cases, federal agencies). Collaborative university-industry teams are incentivized through the grant process, and increasing the level collaborative relationships around data sciences problems is one of the model objectives.

- **Industry Sponsored Applied Research Group** – Uses an industry consortium model to establish an applied research group funded through member organizations. The applied research group functions as an independent entity that addresses specific industry member needs using a dedicated staff and can be a for-profit or nonprofit organization. IP sharing and joint governance models are typically incorporated into the consortium to determine strategic priorities for solutions development, and university faculty are typically only involved in funded consultant roles.

Table 24 outlines additional key attributes of these different types of models for comparison.

Table 24. Key Attributes and Examples of Public-Private Collaborative Engagement Models in Data sciences

Model Type	Industry Pull	Technology Push	Faculty Access	Talent Pipelines	Consulting Services	Networking	Education Programs	Funding Model	Example of Model
Industry Affiliates Portal/Single Access Point	✔		✔	✔	✔		✔	Seed grants, engagement fees	Columbia University Data Science Institute Industry Affiliates Program
Embedded Applied Research Company/ Proximity		✔	✔	✔				Company revenue	Applied Brain Research (ABR) and University of Waterloo
Industry Membership/ Advisory Board	✔		✔	✔	✔	✔	✔	Membership funding	MIT Institute for Data, Systems, and Society
Centralized Innovation Hub	✔	✔	✔	✔		✔		Sponsorships, grants	RiseLab at UC Berkeley
Data Science Environment (DSE)		✔		✔	✔	✔	✔	Grants, sponsorship/ foundation/ institutional funds	Moore-Sloan Initiative, NYU, UC Berkeley, University of Washington
Grand Challenge Projects		✔	✔			✔		Grant awards	NSF Big Data Regional Innovation Hubs (BDHubs) initiative, Michigan Institute for Data Science (MIDAS)
Industry Sponsored Applied Research Group	✔				✔	✔		Membership funding, revenues	Columbus Collaboratory

Previous and ongoing collaboration initiatives in Indiana also present validated examples of ways in which in-state industry and universities have formed effective partnerships that might be used as a model for structuring interactions in AI-related efforts. Figure 19 below outlines several of these mechanisms and the ways in which they have advanced capacity building in specific technology areas or markets.

Figure 19. Examples of Indiana-Based Collaborative Engagement Models in Key Technology Areas

INDIANA CTSI Clinical and Translational Sciences Institute

- The Indiana CTSI, founded in 2008 and funded via a competitive NIH award, is a statewide research partnership/institute among IU, Purdue, and Notre Dame working to solve the State's most pressing health challenges.
- Since its inception, CTSI has secured more than $88M in NIH research funding which has built research infrastructure, attracted top talent, and worked to improve the greatest health challenges.

INDIANA BIOSCIENCES RESEARCH INSTITUTE

- IBRI, founded in 2012, is an independent, nonprofit discovery science and applied research institute currently targeting diabetes, metabolic disease, poor nutrition and related health data science.
- The IBRI exists to bring together companies and universities to work collaboratively on interrelated health issues that are of both global significance and have a disproportionate impact on Hoosiers—diabetes, metabolic disease and poor nutrition.

INDIANA INNOVATION INSTITUTE

- IN3, launched in 2017, is an applied research institute established to solve real-world problems impacting the DoD and Industry.
- Brings together Government, Industry & Academia in cutting-edge technology (e.g. Trusted Microelectronics).

Indiana's current ecosystem conditions, as well as the need to address significant demand from industry, likely require a combination of several of the various types of models outlined above in order to best foster growth in AI-related technology and talent development. A successful model will focus on those engagements that are meaningful and productive for both industry and universities. Companies will have access to leading researchers who are at the forefront in the field in a manner that makes "poaching" unnecessary and unappealing while the university will have opportunities to generate meaningful research with new funding sources and the ability to impact students. Elements of reliable access point, embedded applied research functions, industry advisory board, and data sciences environments are all relevant to particular aspects of Indiana's challenges today, and both industry and university partners will need to embrace a hybridized model that can address key issues and elevate the state's profile in AI-related capabilities.

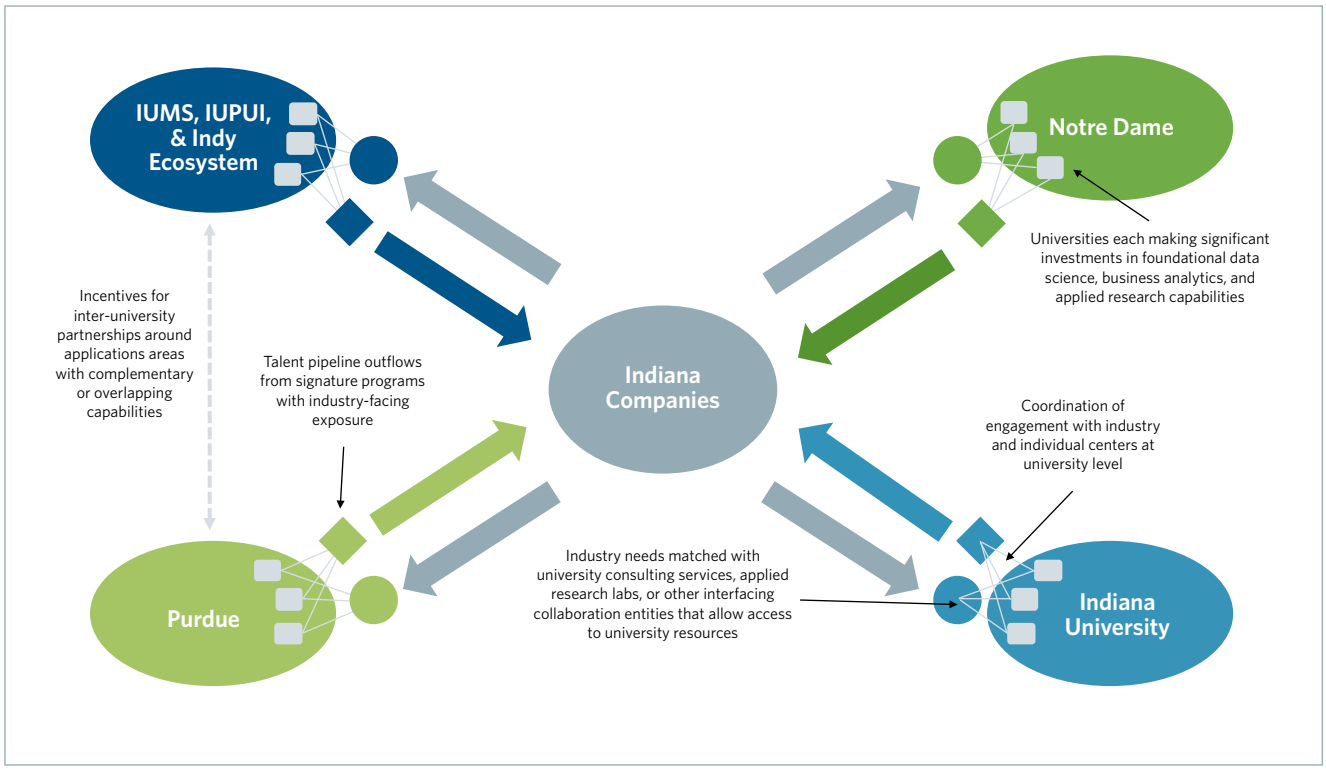
C. Strategic Recommendations for Growing an Indiana AI-Enabling Ecosystem

Goal of Strategic Recommendations: Create an Active, Sustainable Community of Practice for Indiana Around Applied AI and Advanced Analytics

Indiana is beginning to increase the level of attention and investment in AI-related capabilities, with a number of recent announcements and initiatives on the part of both industry and university leaders that are designed to directly impact the state's position. Indeed, despite the challenges noted above there are a number of factors aligning for the state that can be collectively leveraged to build momentum towards a more competitive position for the state: major companies and their leadership are engaged in serious strategic implementation efforts around data sciences; there is an emerging base of specialized entrepreneurial companies who are providing services in this space; there is a significant base of university R&D and talent expertise coupled with a willingness to engage with industry partners, and a robust volume of skilled talent being created from regional pipelines with modern data sciences skill sets.

In order to create an AI-enabling ecosystem, Indiana must connect and coordinate its assets to grow a broader community of excellence, which in turn can attract and retain business and talent in growing applications areas. And, this engagement will need to be activated with intentional action connecting university and corporate talent on real-world problems. The Innovation District at 16 Tech was built to enable this activity and connectivity and should be the center of the initial engagement. To be successful over the long term, the structure of this connectivity has to be adapted such that the needs of both universities and companies are met rather than fitting a model that conflicts with strategic priorities of the stakeholders in the ecosystem. Growing the community of practice must meet a university need for a model that does not force them to work against their own strategic investments in this space, an industry need for short term solutions that are tailored to highly specific engagements, and a talent need for an in-state environment that is attractive and growing. Once engagements have begun and matured over time, Indiana could be home to an ecosystem where corporate and university engagements that drive innovation routinely occur as a matter of course. Figure 20 outlines an aspirational view of this model, with utilizes elements of various public-private collaboration structures in data sciences found throughout the country.

Figure 20 An Aspirational Vision for a Robust AI-Enabling Ecosystem in Indiana

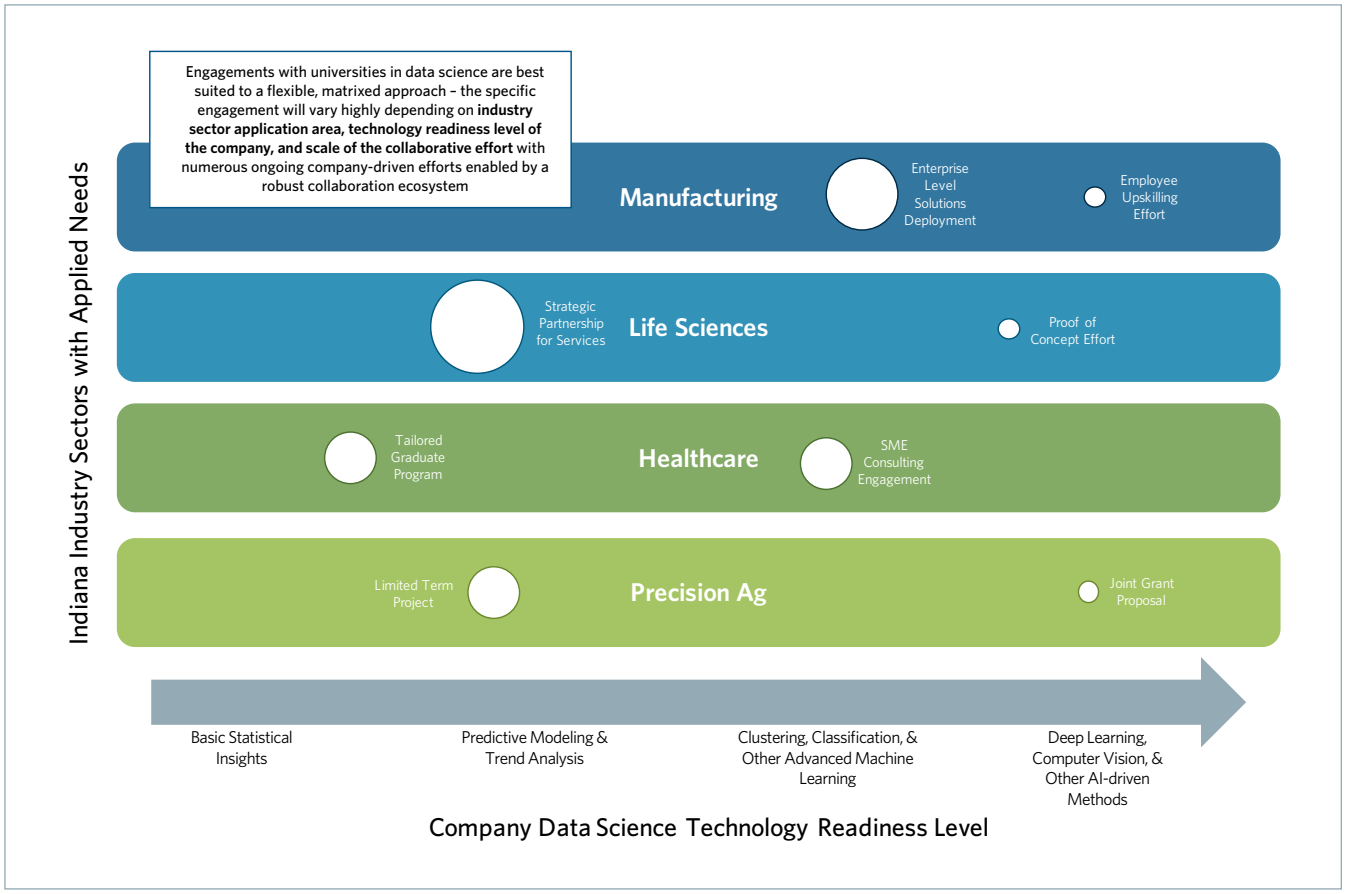


TEConomy concludes that engagements with Indiana universities are required across two dimensions:

- A flexible, matrixed approach that is adaptable to the fact that specific engagements will vary highly depending on industry sector application area, technology readiness level of the company, and scale of the collaborative effort due to the diverse nature of Indiana industry stakeholders. Development of broader solutions at the industry sector level are unlikely to meet the specific needs of the majority of companies in that sector due to a high level of variance in technological readiness and business norms (with the exception of healthcare analytics as noted below).
- Development of a joint university resource hub in Indianapolis that can serve to interface with companies and provide a robust access point for linking companies to the analytics resources of Indiana universities (in R&D and in education) whether those resources be in Indianapolis or at other in-state campuses. The resource hub would also be a potential home to the development of a Indianapolis-based healthcare analytics collaboration building upon the significant base of assets identified.

Figure 21 outlines the matrixed approach to structuring engagement at the company level, with the goal of creating sustainable engagements with companies that can prove short-term value and progress to large, enterprise level scales of engagement over time. **This type of engagement also builds a strong network of numerous inter-institutional and public-private relationships, which is then highly resilient and adaptable to market changes in ways that more siloed engagements are not.**

Figure 21. Matrixed Approach to Structuring University Engagement with Indiana Companies



Realigning talent flows also requires more active relationship-building on the part of industry, which more coordinated relationships can help activate. Indiana's skilled graduates with AI-related backgrounds are typically being attracted away in a competitive market rather than choosing to leave due to perception of a lack of opportunity, meaning a key aspect of coordinating relationship building will be increasing awareness of in-state opportunities for meaningful careers and better communicating the value proposition of the state's brand. Over time, the creation of a highly innovative community of practice based around regional in-state hubs will build centers of gravity that retain cohorts of talent with varying specializations. Significantly boosting retention in the short term, however, may require significant use of incentives and active marketing efforts on the part of industry stakeholders with sizeable needs.

Specific Recommendations for Advancing Indiana's AI-Enabling Ecosystem

In order to address the identified industry needs and challenges and work towards creating a high-functioning ecosystem in AI-enabling activities, several specific activities are recommended across three thematic areas. These areas include:

- Coordination of Industry Engagement and Continuing to Attract Top Research Talent at Indiana's Research Universities,
- Addressing Demand for Talent and Continuing Education of Existing Workforce,
- Coordinating Access and Collaborations in University-Industry Partnerships and Talent Access in Indianapolis, and
- Enhancement of Indiana's Profile in AI.

1. Coordination of Industry Engagement and Continuing to Attract Top Research Talent at Indiana's Research Universities

The overarching need for Indiana revolves around intentional engagement where intellectual talent from university and corporations work and interact in both structured and unstructured ways.

- Support and enable traditional sponsored research by taking a more proactive approach to understand data sciences needs of Indiana companies and then making connections to relevant university capacities. Use engagement offices that remain knowledgeable about the comprehensive profile of ongoing university activities in AI-related technologies and facilitating industry relationships with the right university partners.

- Fund and develop applied projects with industry partners where a question of interest is studied by a team made up of individuals with diverse experiences and expertise from academic and corporate institutions working together in a set location for a fixed period of time. The final product could be a presentation, publication, or spin-out. Cohorts working together on such projects necessarily build longer lasting relationships that build community and allow for future opportunities for engagement and problem solving. University personnel can participate through competitive fellowships or sabbaticals.

- Encourage research universities in the state to further adopt interdisciplinary institute models for advancing data sciences within their programs to help aggregate activity institution-wide. Build methodology expertise with recruitment of new talent in mathematical and computations areas that bridge to health and life sciences, manufacturing, and other areas where Indiana has concentrated industry presence. Concentrate on identifying and attracting talent with National Academy or Lasker award winners, for example, or those rising stars currently working at recognized centers at Stanford, MIT, Northwestern, or Toronto. Establish competitive fellowships, à la MacArthur Foundation for existing faculty to enable to pursue new research directions that build on expertise in data sciences. Look to refine Indiana's data assets and technologies and invest in the creation of "data lakes" that can be used to drive innovation in areas of particular concentration and strength such as healthcare delivery and manufacturing. Use AI tools to concentrate data to drive simulations, analyze the past, optimize the present, predict the future, test hypotheses, and explain inferences.

2. Addressing Demand for Talent and Continuing Education of Existing Workforce

Addressing the significant needs of Indiana's industries for skilled talent in data sciences and analytics forms the foundation for growing the state's AI-related capabilities, and primarily involves retaining more graduates from the significant talent pipeline that exist in-state as well as continuing education of existing workforce.

- Establish funding programs for data sciences environments (DSEs) based on industry partner, live-learn community model to bolster pipeline and build student awareness of Indiana opportunities. An initiative in this area would ideally encourage establishment of these programs at all major research universities which could then be connected with the coordination function described above to engage industry partners. Purdue's Data Mine is an example of a developing world class DSE program that is organized around industry engagement and immersive skills-building in data sciences that can serve as a model for other universities.

- Implement incentives programs to retain skilled analytics talent in-state upon graduation. Options to explore include tenure-conditional "signing bonus" awards, grants for ongoing education, tuition reimbursement, state tax rebates, and other funding incentives in targeted programs or skill sets aligned with data sciences needs of industry stakeholders.

- Significantly increase the level of in-state industry marketing and presence at hub locations to engage Indiana talent pipelines. Industry stakeholders can pursue more aggressive recruitment strategies in conjunction with coordination efforts described above.

- Maintain comprehensive information on graduates in relevant program areas who leave the state for their first employment position in order to pursue recruitment for senior roles in the future.

- Implement new and expand existing continuing education programs for workforce at Indiana companies. Expand engagement to include visiting lecturers from companies to provide living lab opportunities for problem solving.

3. Coordinating Access and Collaborations in University-Industry Partnerships and Talent Access in Indianapolis

Indianapolis has a significant presence of academic assets and programs of degree study emphasizing advanced analytics – particularly in, but not limited to, healthcare applications and informatics. The IU/Purdue/IUPUI academic constellation in Indianapolis comprises biomedical and health sciences, informatics, computer science, engineering and associated degree programs of IU and Purdue, with activity centrally concentrated at the IUPUI campus in Indianapolis. With Indianapolis being home to significant major corporations (as well as midsize and entrepreneurial growth companies) and major healthcare institutions, there is a built-in regional demand for engagement between these employers who need to apply advanced analytics and the local presence of the universities which can form hubs for talent supply, and regional access points for statewide university capabilities in research, education, training and consultative services.

The fast growth and evolution of careers requiring analytics competencies and credentials, means that it will be increasingly important for education and ongoing certification and recertification courses to be readily accessible to regional employers. A joint university resource hub in Indianapolis that can serve to interface with companies and provide a robust access point for linking companies to the analytics resources of the universities (in R&D and in education) whether those resources be in Indianapolis or at other in-state campuses, would also ensure robust communication of industry needs and projections for talent and skills demands into the future. The resource hub can make the appropriate connections with companies to Indianapolis-based faculty, or to faculty located at other campuses, and provide a core venue for ongoing discussions of talent development and talent retraining program development.

4. Enhancement of Indiana's Profile in AI

For Indiana to establish itself as a credible leader and destination location for the development and deployment of leading technologies relating to artificial intelligence, internet of things and advanced data analytics, it must enhance its profile.

- Establish a signature statewide data sciences conference with a rotating in-state university host and awareness-building branding around state industry opportunities and leading applied research activities. Leading companies should be identified to serve as potential sponsors and participants to build knowledge of in-state employment and collaboration opportunities.
- Leverage 16 Tech, an innovation community, as a multi-institutional hub located in the Indianapolis environment focused around healthcare applications of AI and analytics that leverages significant institutional and industry strengths in central Indiana.
- Incentivize entrepreneurial industry-university partnerships in data sciences by mitigating disincentives for faculty to participate in commercialization activities:

    - Support data sciences spin-out companies founded by faculty-entrepreneur partnerships using competitive grant award model and prioritizing companies developed around data sciences services that can be leveraged by larger industry stakeholders; and
    - Explore funding and incentives for faculty involved in industry consulting and research, including career advancement credit and flexible schedule structures.
- Coordinate state institutions around common mission and complementary areas of specialization to enable competition at scale with other leading ecosystems. In particular, encouraging institutions to reach consensus on areas of specialization in applied analytics and AI to avoid duplicative investment and establishing protocols to reduce in-state competition for talent and resources is critical to long-term success.

D. Conclusion

Indiana's economic future is and will continue to be impacted by the role of AI and advanced analytics in innovation. Its corporations and universities have made substantial and strategic investments in expanding, improving and building capabilities. Indiana's ability to compete will depend upon its industrial and academic sectors to drive more collaborative engagement with each other in order to grow and maintain an ecosystem that can attract and retain talent while supporting the growth. There is a significant pipeline of advanced analytics talent in the state and it is being further expanded through focused investments and innovative program expansions at the major research universities in the state. Furthermore, the universities are structuring transdisciplinary analytics initiatives and consulting access points to enable faculty and other university research personnel to interface with companies on joint R&D projects and specific challenge engagements.

Because of high competition for analytics talent, and aggressive recruitment of Indiana talent by out-of-state enterprises, it is important that Indiana's advanced industries be proactive in interfacing with universities to build early relationships with analytics student populations. It will also be particularly important for industry to engage in intensive internship engagements with students, often across multiple years. It is likely, however, that beyond increasing opportunities for connectivity, Indiana industry will need to raise the salaries it offers for analytics positions to remain competitive for attraction and retention of talent.

Indiana has a significant opportunity to advance its competitive position in the AI-related technologies and talent space through coordinated efforts to connect and realign stakeholder relationships within its ecosystem. The state's path to success lies in leveraging these technologies towards leading applications in its major state industry clusters rather than in pursuing research enterprises in more fundamental "basic data sciences."

Appendix

Innovation Landscape Network of U.S. AI Companies Receiving Significant VC Investment, 2014-2018

Source: TEConomy's network analysis of PitchBook Venture Capital Database firms. Note: the analysis is limited to U.S. companies focused on AI-related technologies and services receiving at least $0.5 million in VC funding during the 5-year period.

Fig A-1. Innovation Landscape Network of US AI-Related Companies Receiving Significant VC Investment

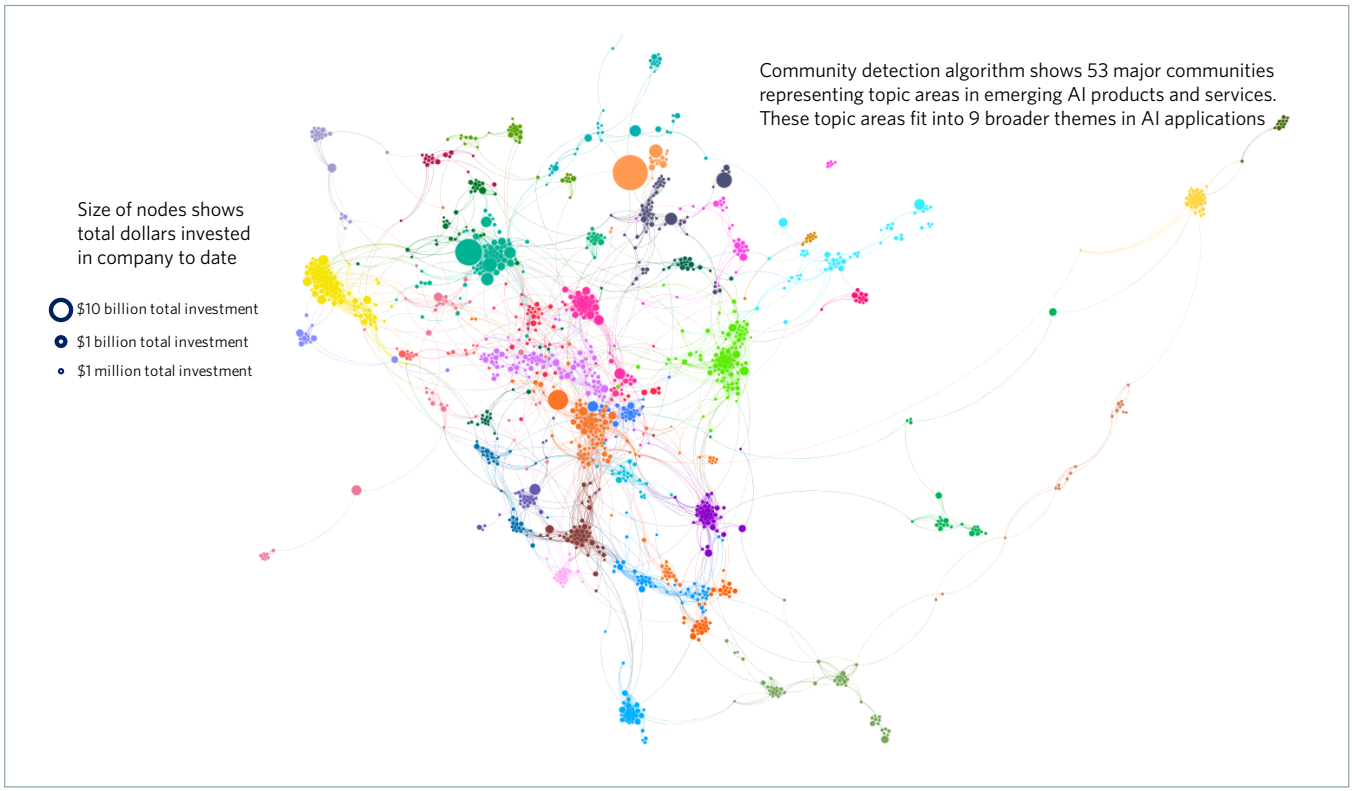


Fig A-2. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Computing and AI Tools

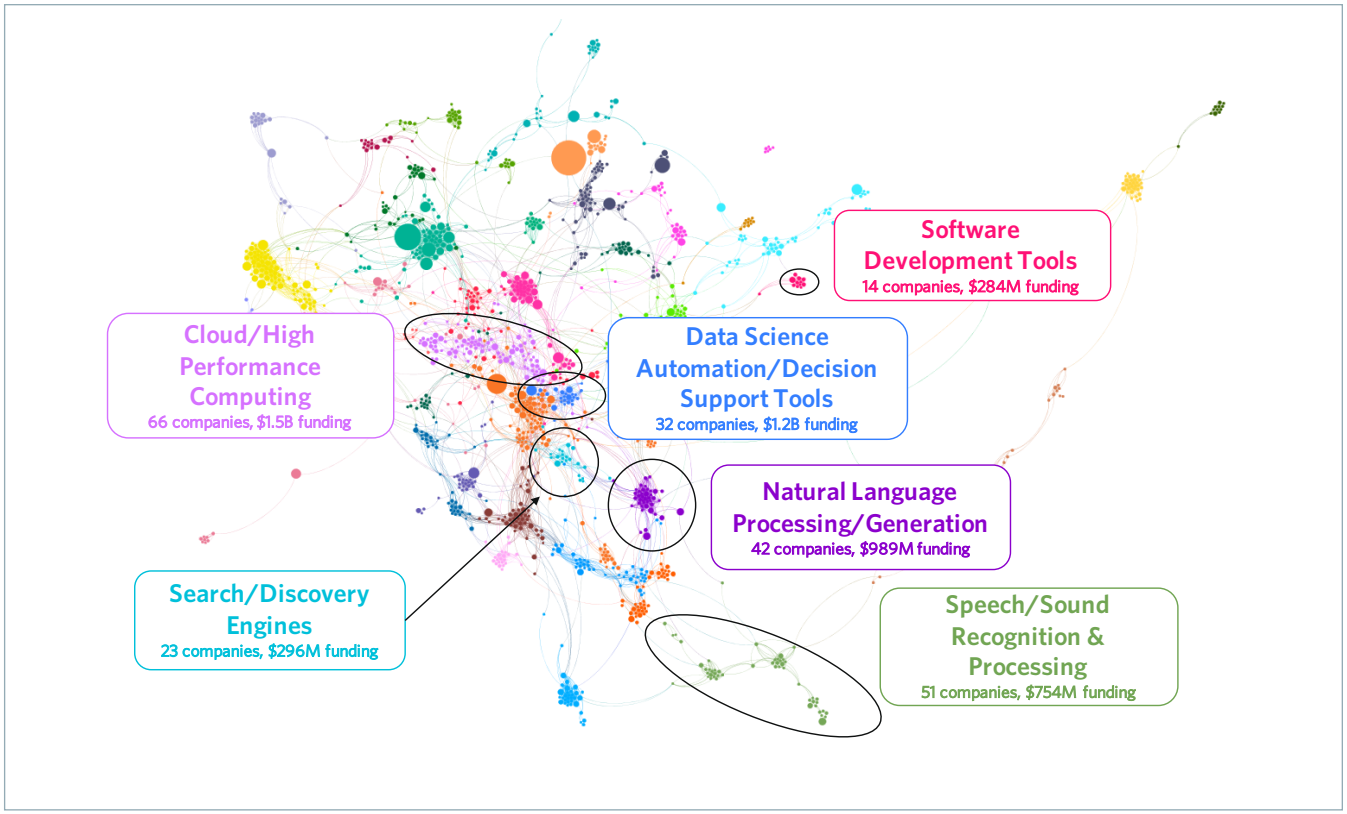


Fig A-3. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Autonomous Systems

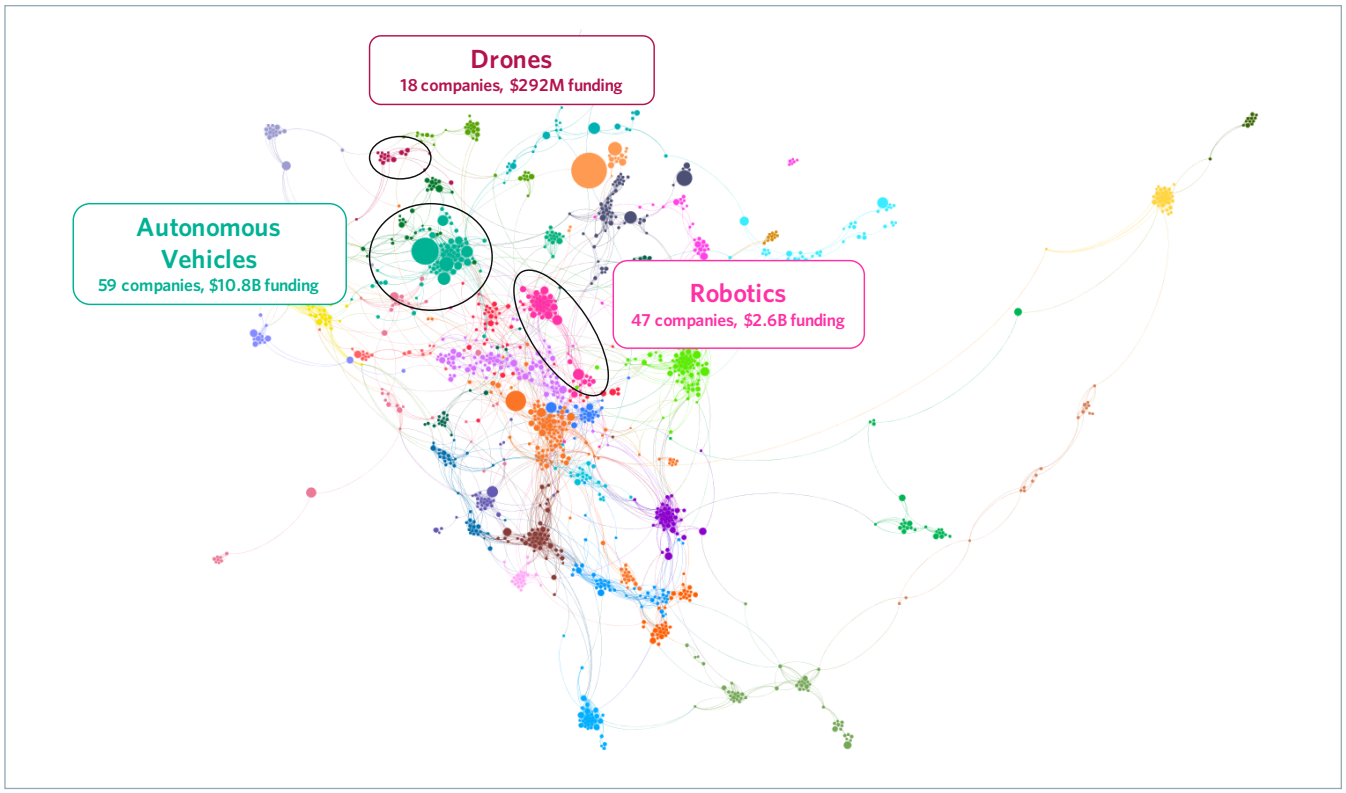


Fig A-4. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Business Services Analytics & Process Automation

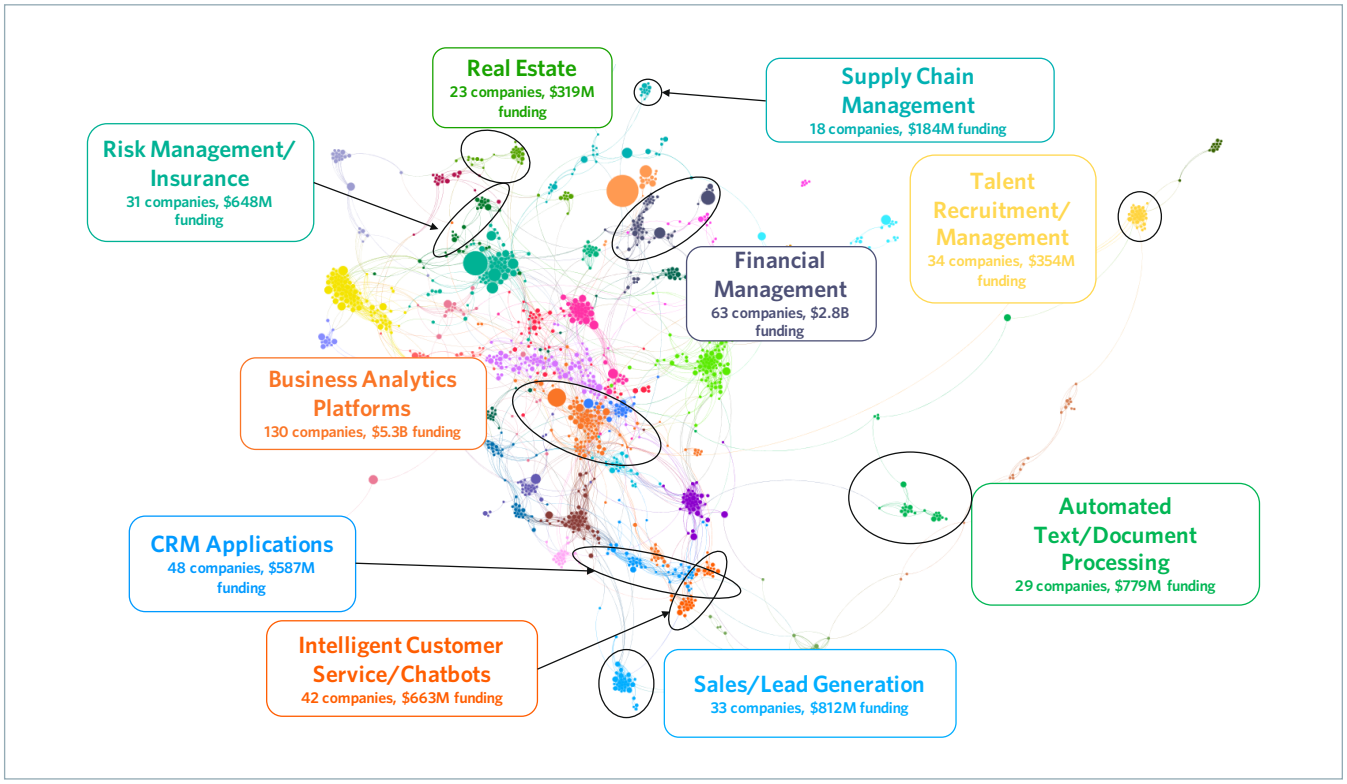


Fig A-5. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Marketing & Content

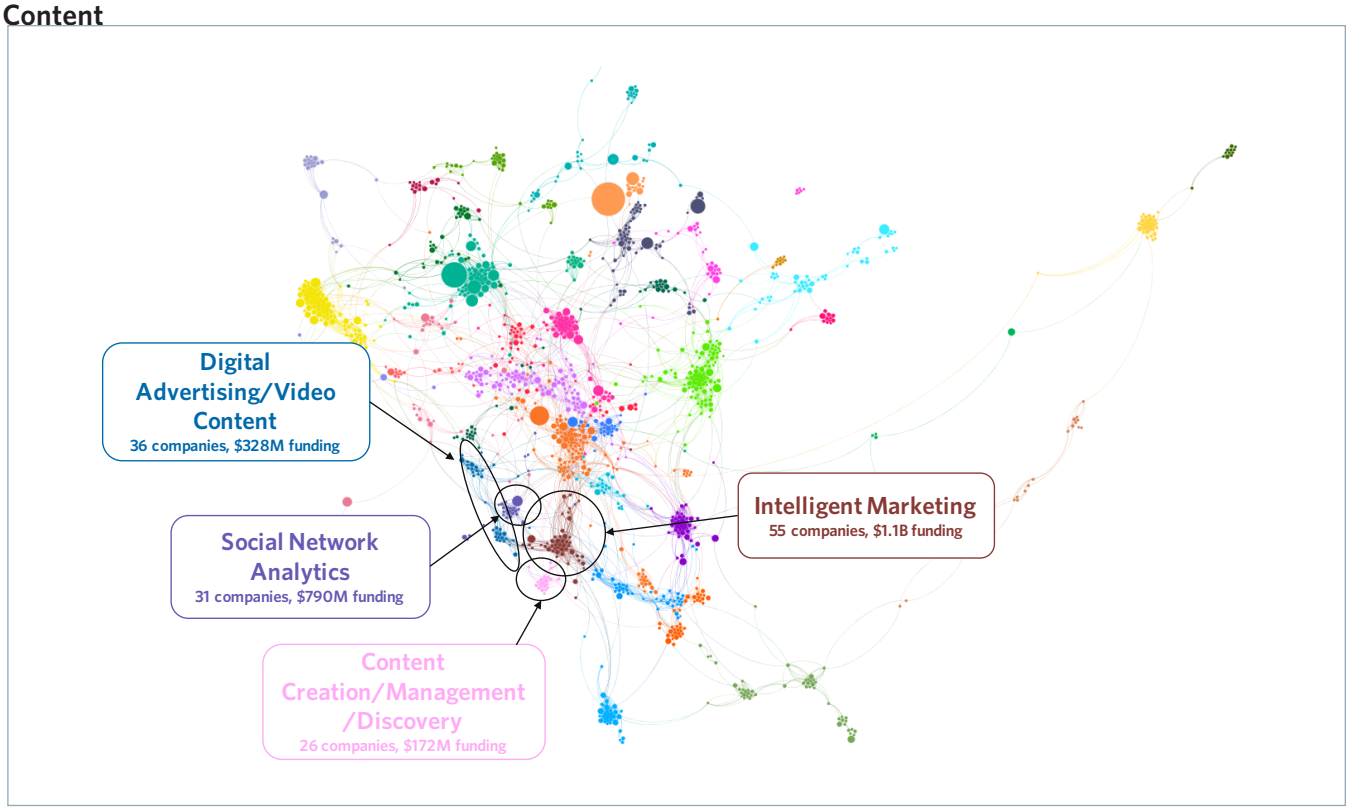


Fig A-6. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Education; Entertainment & Consumer Goods/Services

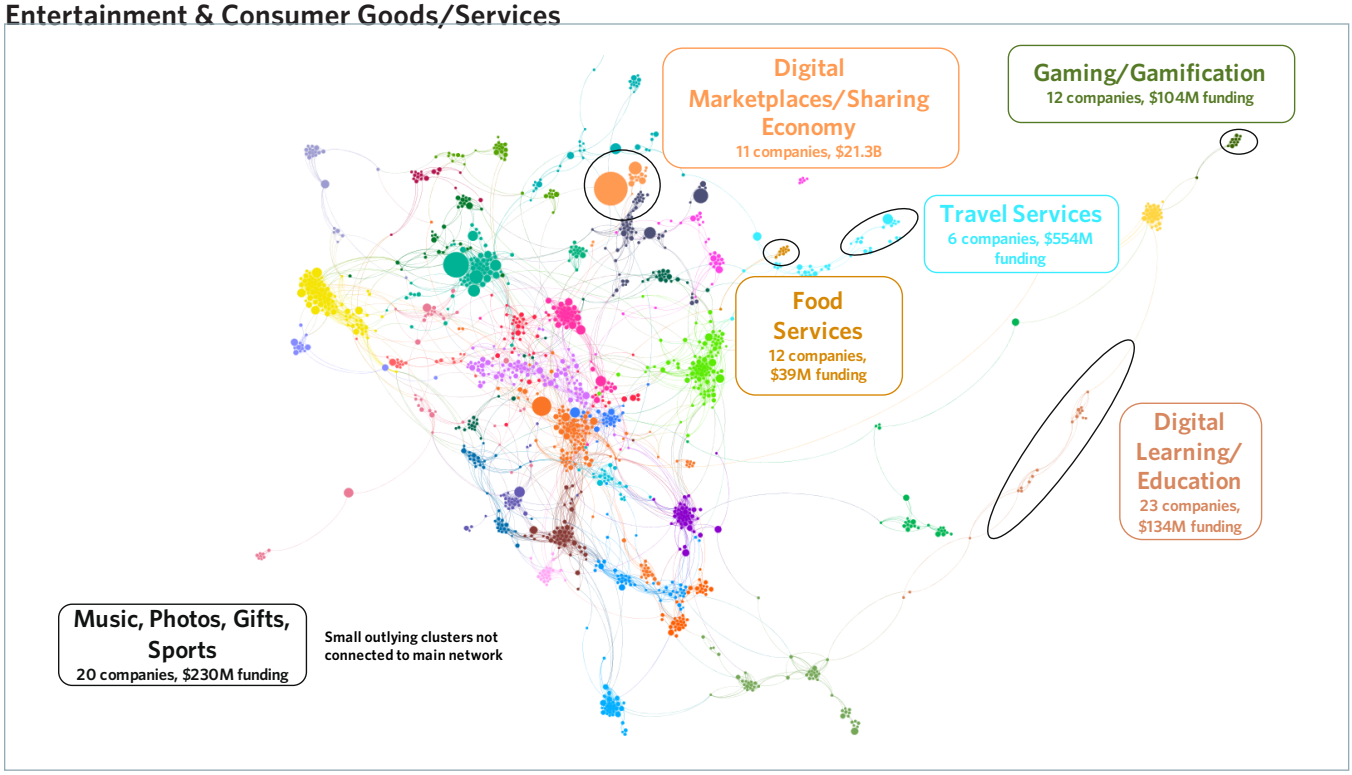


Fig A-7. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Biomedical

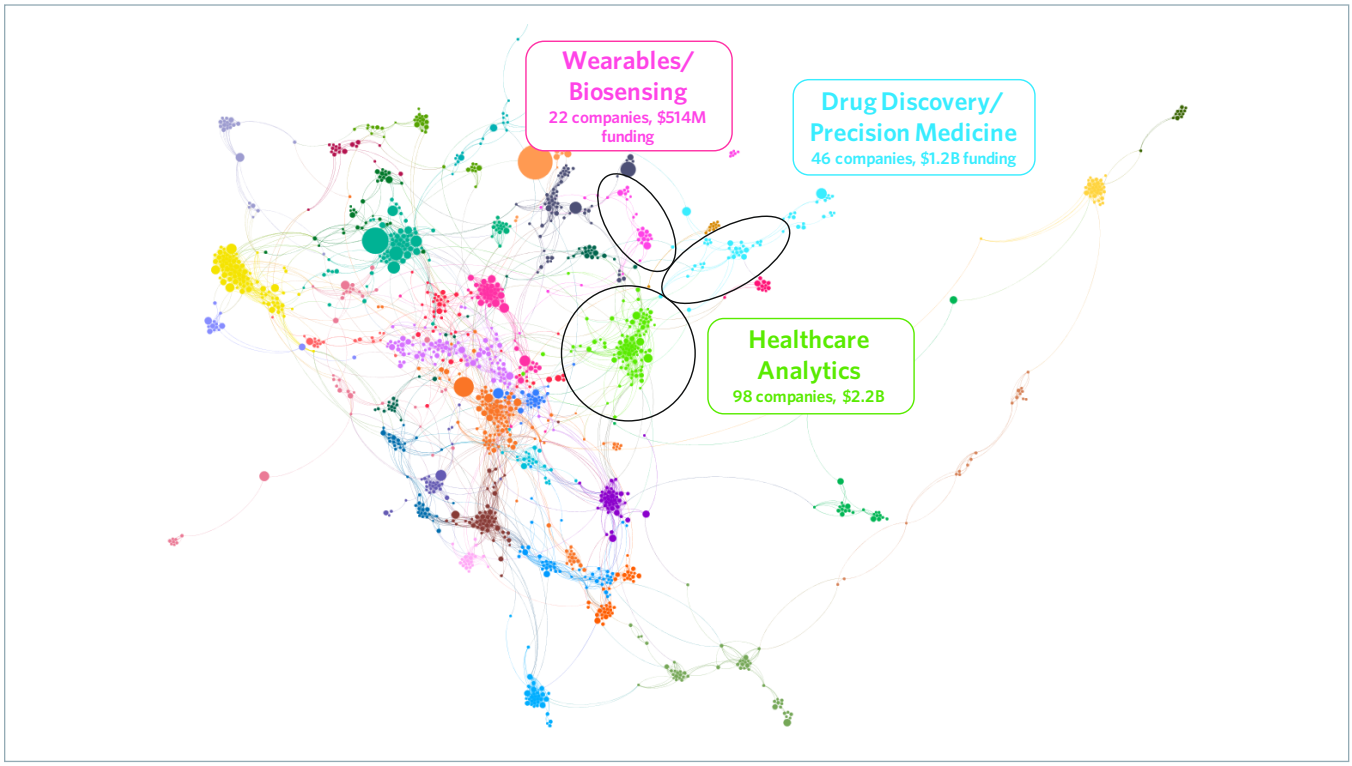


Fig A-8. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Imaging

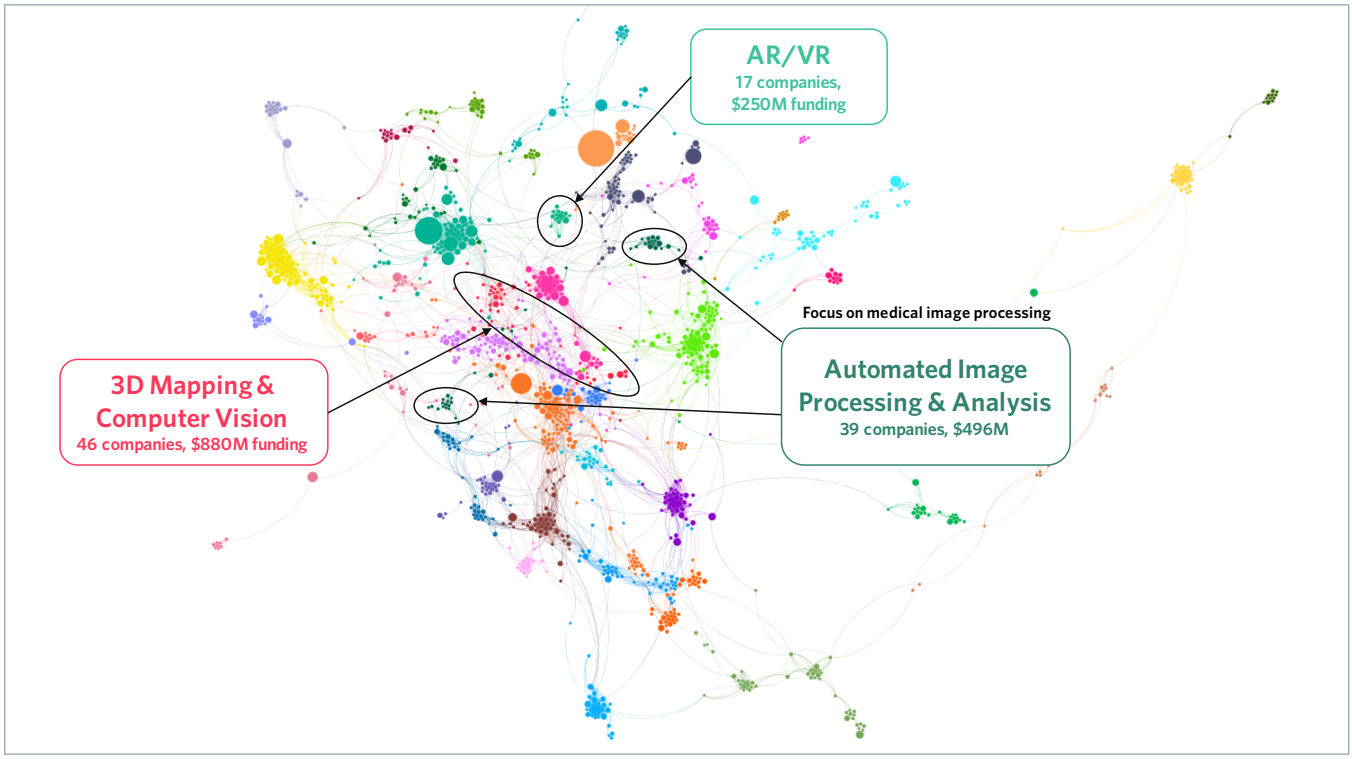


Fig A-9. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Infrastructure, Industrial Controls, & Environment

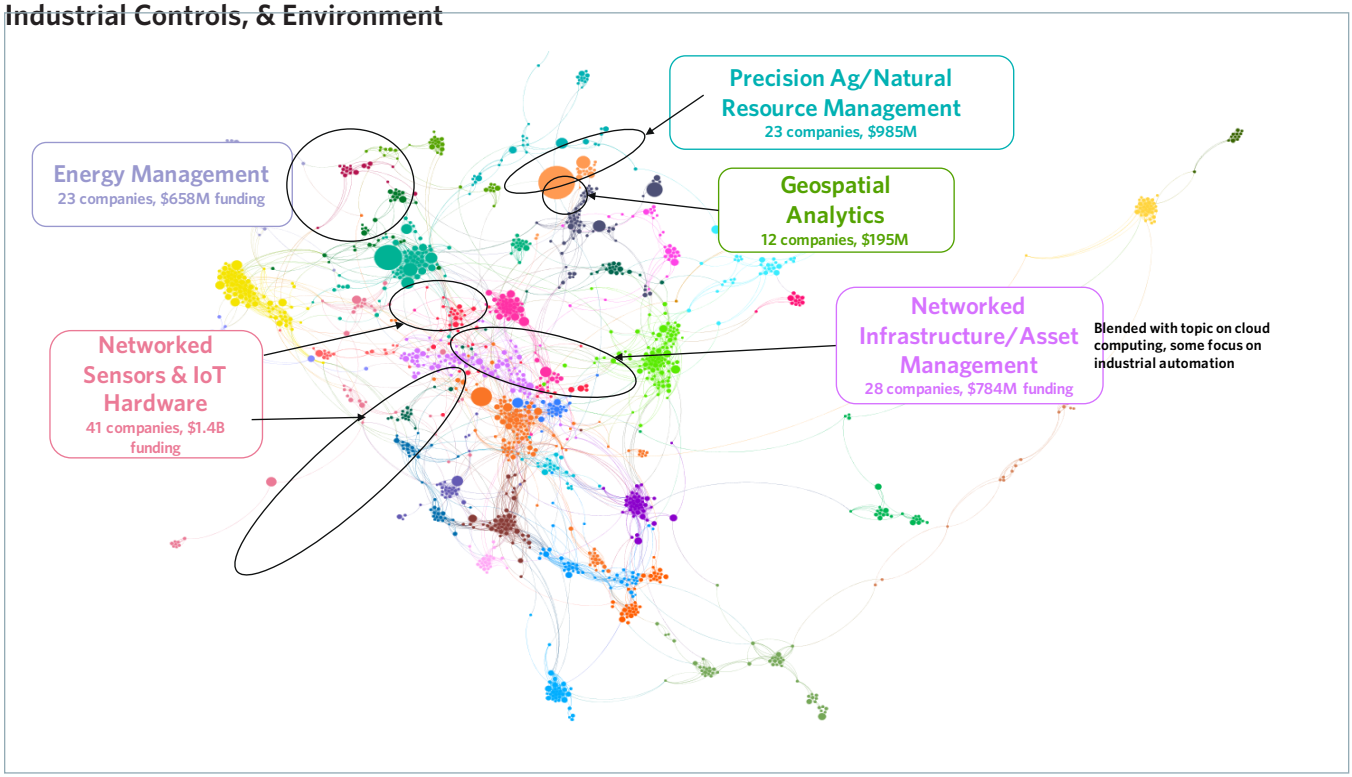
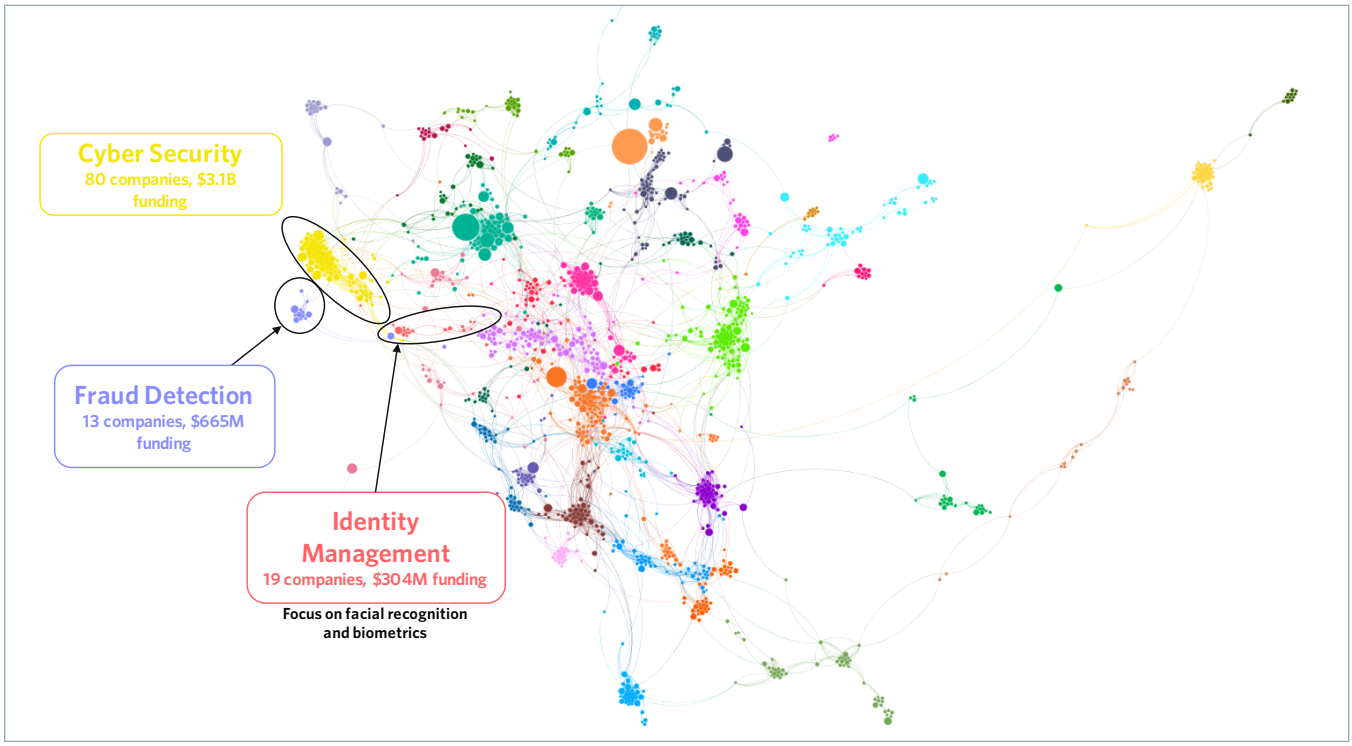


Fig A-10. Broad Theme in Innovation Landscape Network of US AI-Related Companies: Security



TECONOMY
PARTNERS LLC