Similarity Report

Original Document:

Telematics and Informatics Reports 14 (2024) 100127
Available online 7 March 2024

Artificial intelligence research: A review on dominant themes, methods, frameworks and future research directions

Kingsley Ofosu-Ampong
a,b
a
Alliance Bioversity International & CIAT, Ghana
b
Heritage Christian University, Ghana

ARTICLE INFO
Keywords:
Artificial intelligence
Classification
Literature review
Technological issues
Research agenda

ABSTRACT
This article presents an analysis of artificial intelligence (AI) in information systems and innovation-related journals to determine the current issues and stock of knowledge in AI literature, research methodology, frameworks, level of analysis and conceptual approaches. By doing this, the article aims to identify research gaps that can guide future investigations. A total of 85 peer-reviewed articles from 2020 to 2023 were used in the analysis. The findings show that extant literature is skewed towards the prevalence of technological issues and highlights the relatively lower focus on other themes, such as contextual knowledge co-creation issues, conceptualisation, and application domains. While there have been increasing technological issues with artificial intelligence, the three identified areas of security concern are data security, model security and network security. Furthermore, the review found that contemporary AI, which continually drives the boundaries of computational capabilities to tackle increasingly intricate decision-making challenges, distinguishes itself from earlier iterations in two primary aspects that significantly affect organisational learning in dealing with AI's potential: autonomy and learnability. This study contributes to AI research by providing insights into current issues, research methodology, level of analysis and conceptual approaches, and AI framework to help identify research gaps for future investigations.

1.Introduction

In recent times, the widespread adoption of computers has surged dramatically, fundamentally altering the dynamics of business operations and competition as an imperative for long-term viability. To address these imperatives, there is a need for efficient allocation of resources and innovations. Artificial Intelligence (AI), an evolving innovation, aims to address these imperatives. While AI is not entirely novel, its commercialisation gained momentum around 2000. AI, in essence, is a simulation of human intelligence in computer systems [1]. AI comprises the development of algorithms, software, and hardware that enable machines to perform tasks that typically require human intelligence [2–4].

AI is still garnering attention, leading to a slow but steadily growing body of research (e.g. [5]). While these reviews have provided few valuable insights into AI in other domains [6,7], huge knowledge gaps persist, underscoring the need for further examination of information systems (IS). Thus, AI in information systems research is a new technology for gathering information, generating results, interpreting it, improving decision systems and being able to interact with its environment [8]. The decision-making process of AI is often not transparent and determinative [9]. Thus, a deeper understanding of AI in IS is needed to establish conventional functions of AI studies.

Furthermore, despite the increasing popularity of AI and its societal influence [10–12], we lack a comprehensive understanding of what we know and what we need to know about AI and how people experience it. Prior studies have called for researchers to give more clarity in defining AI, even if that means redefining it away from traditional human intelligence and conducting more research on societal and personal effects that advance AI in everyday life [13]. This study seeks to provide an enhanced understanding of the knowledge and conceptual gaps and contribute to the practical implementation of AI in IS. Concisely, this paper provides a review and analysis of artificial computing from 2020 to 2023. The emphasis is on dominant theories and themes, methodologies, frameworks, trends and research direction for understanding AI in recent times.

1.1. Early days and initial paradigms of artificial intelligence

There is an extended history of AI, but its modern iteration evolved

E-mail address: k.ofosu-ampong@cgiar.org.

Contents lists available at ScienceDirect
Telematics and Informatics Reports
journal homepage: www.elsevier.com/locate/teler
https://doi.org/10.1016/j.teler.2024.100127

around the 1950s. Credit to Alan Turing and the conference held at Dartmouth College, the term Artificial Intelligence was framed and defined as “the science that makes machines intelligent” by John McCarthy in 1956 [14,15]. Thus, the early AI focused on machine development that was capable of making decisions that only humans could accomplish. The initial paradigm of AI had the potential to reason multi-step, create innovative solutions, comprehend natural language and even contemplate its own thinking – hereby referred to as strong AI [13]. The main characteristic of the strong AI is its general symbolic manipulators (reasoning); however, it lacks the progressive capabilities of the 21st-century AI model.

From 1974 to the early 1980s, a period referred to as the first winter of AI witnessed reduced funding. Consequently, it lowers the hype and interest in the field of AI [16]. Fast forward to 1997, IBM’s Deep Blue computers surpassed the reigning world chess champion – G. Kasparov. This defeat is important in the history of AI research because it proved that computers could surpass human intelligence in goal orientation. Subsequently, deep blue became an integral part of financial modelling and health systems resilience.

The early 2000s to date started to witness Stanford vehicles driving autonomously across the desert, while IBM’s Watson won the “Jeepardy”. Watson’s feat was essential to the development of AI breakthroughs in Natural Language Processing. This breakthrough helped to dispel fears and some scepticism surrounding AI and highlighted numerous ways of benefiting humanity. Generally, the idea of AI is ancient until its significant technological advancement in the mid-20th century [15]. However, the explosion of ubiquitous AI and real-world applications is relatively new and rapidly evolving, hence the conceptualisation of divergent perspectives.

1.2.Definition of artificial intelligence

The AI phenomenon has its genesis in other technologies, namely computing power, machine learning, big data, cloud computing, open-source software, algorithms and virtualization [17,18]. Features that distinguish AI from related technologies are its ability to learn, adapt, reason and decision-making. In addition to handlinguncertainty and noise, interaction and collaboration, and representation and abstraction [19,20]. Currently, there is no standard definition of AI. However, significant strides are made by academics, government institutions and industry players for a standard definition of AI components. The lack of standardization stems from the fact that smart technologies such as smartphones and smart homes are referred to as AI just as autonomous technologies (e.g. self-driving cars and drones) [21]. Due to the lack of a standard definition of AI, current literature reviews call for studies and reviews to broaden the scope of frameworks, themes and research direction. However, an attempt by [11] define AI as an “unnatural object or entity that possesses the ability and capacity to meet or exceed the requirements of the task it is assigned when considering cultural and demographic circumstances.” For this review, AI is defined as a model that possesses the ability to reason, learn and act autonomously as human behaviour [22]. Recent literature reviews [10,11,13] on information systems in AI have tried to address the definitions, contextualisation, business value and frameworks of AI; however, several research gaps exist to necessitate this study (see Table 1).

1.3.Domain terminologies in artificial intelligence

Researchers have often described AI with numerous domain technologies that range from natural language processing (NLP) to computer vision and machine learning. While these terms are closely related to artificial intelligence, they differ in their goals and dedicated purpose. Some articles used AI, NLP, computer vision, machine learning, autonomous robotics and recommender systems interchangeably and did not clearly define them. Based on the frequently used terms in the review, the following are clearly defined:

Natural language Processing describes the interaction between computers and human language and includes technologies like language transactions, chatbots, sentiment analysis (emotional tone of text) and entity recognition to function [24].

Computer vision constitutes an enabling machine used to understand and interpret visual data from videos and images. Some of the key features include object detection, image classification and segmentation, and facial recognition.

Machine learning is a foundational technology of AI and includes supervised learning, unsupervised learning, deep learning and reinforcement learning (training agents) necessary to initiate communication. Compared to supervised learning, unsupervised learning tends to discover patterns in an unlabelled data structure.

Autonomous robotics describes how machines in real-world environments operate independently when AI is integrated with robotics. Key enabling technologies of autonomous robotics include sensor fusion, simultaneous localization and mapping, and path planning and control.

Recommender systems consist of collaborative filtering, content-based filtering and hybrid approaches that are used to provide personalized recommendations to users.

2.Framework taxonomy

Artificial intelligence is a rapidly evolving, complex and diverse field with various applications and components, making the classification of

its literature a challenge. Given its wide-ranging applications and
diverse subfields, any attempt to categorize AI literature must consider
the many facets and complexities inherent to this rapidly evolving field.

Table 1
Research gaps in prior reviews.

Article Research issue Conceptual
approaches
Identified gaps
[10] ü Business models
of AI
ü Bibliometric
analysis and
review of AI
ü Role of AI in
sustainable
business models
Cultural drift based
on AI, sustainable
business model and
knowledge
management
systems
Business model
concepts,
organisational
memory, AI
methods and
networks
ü Need to adopt a
quantitative approach
to review AI research
ü More systematic
review for our
understanding of AI in
business
[13] ü Review on the
business value of
AI
ü AI in information
systems
implications
Value types of AI
include:
i) process
automation,
cognitive
engagement and
cognitive insight
Conversational
functions typical of
the human mind
ü Lack of consensus
around the definition
and themes of AI
ü Need for review on
technological,
performance, and
contextual aspects of

AI
[23] ü How to effectively integrate AI and organizational strategy
Conceptual approaches involve the intersection of AI themes, sources of value creation and business strategy themes e.g. alignment with IT
ü Insight into new AI tools to align with business strategy and contextual needs of organisation
[11, 22]
ü Review on frameworks and factors influencing acceptance of AI
Framework to assess AI: i) built on adoption theories and actual use behaviour of AI
ii) Issues with AI replication
iii) Reliance on self-reported data
ü Need to examine the impact of AI on the level of analysis
ü Inconsistency in the definition and operationalisation of themes of AI

In this regard, the classification themes of [25]were adapted to reflect the diverse themes in AI literature and its current happenings. Table 2 shows the classification of AI into four (4) main themes and twenty-five (25) sub-themes. The main themes include technological issues, contextual issues, conceptualisation and domains and applications.
The technological issues border on the technological implications of AI. The sub-themes under technology issues consist of data quality and quantity, bias, model interpretability, scalability, ethical AI design, explainable AI (XAI), AI governance and regulation. The contextual issues theme espouses studies that focus on contextual conditions and consequences of AI including sub-themes such as ethical considerations, privacy concerns, bias and fairness, job displacement, social acceptance,

and regulatory and legal challenges. The sub-themes mostly border on studies related to privacy, bias and fairness of AI. Concerning the conceptualisation of AI, the sub-themes include articles that offer insight into AI frameworks and predictability and also its safety and control. AI conceptualisation aims to offer a better understanding and insight into the phenomenon based on contextual development and use. Lastly, the domain and application of AI are focused on sub-themes such as e-health, fraud detection, chatbots, computer vision, game design and personalized learning, as well as smart grids.

3.Methodology

The review primarily relied on electronic database searches, a standard approach in IS research. Initially, the emphasis was on senior IS journals, but these journals yielded few studies on AI (e.g. ChatGPT), likely due to the limited access to papers and conceptual complexity of the subject in recent times. Thus, more articles and publications were found in the Scopus, AISeLibrary and Web of Science databases. Consequently, a more extensive search was conducted across these databases, targeting peer-reviewed scholarly journal articles published between 2020 and 2023. Table 3 shows the strings considered in the search process that yielded 85 articles for analysis.

To ensure the quality of the selected articles, a manual filtering process excluded editorials, review articles, reports, conference papers, dissertations, books, working papers, and articles from non-IS disciplines. Exclusion criteria also applied to non-peer-reviewed papers, duplicate articles, simulation studies and studies with no focus on AI. This meticulous process led to the inclusion of 85 articles for further analysis. Fig. 1 shows the breakdown of the final inclusion criteria. These articles were categorized based on multiple criteria, encompassing themes (comprising four major and 26 sub-themes), research methodologies (qualitative, quantitative, mixed methods, simulation, experiment, or "no method"), geographical focus (continental), level of analysis (micro, meso, macro), publishing outlets (journals), and research frameworks (including a category for studies lacking a specific framework).

In summary, this review leveraged electronic database searches to identify peer-reviewed papers on AI between 2020 and 2023. These articles were subsequently organized and analysed using diverse criteria, providing a comprehensive perspective on the research landscape within this field. The chosen journals for the review are significant in AI research as many strides have been made in defining and contextualising AI. Some of the information systems included in this review include Decision Support Systems, Journal of Strategic Information Systems, MISQ, European Journal of Information Systems, Intelligence, Computers in Human Behaviours, International Journal of Information Management and Information Systems Management.

4.Findings and discussion

4.1.Outlets of publication

The distribution of the journals with their corresponding statistics is as follows. The journal with the highest number of publications among the listed articles is "Nature Machine Intelligence," with eleven articles published in different years (2020, 2022, and 2023). The journal is specifically tailored towards AI and machine intelligence hence no surprise with its appreciable figure (e.g. [6,18,26]). The AI & Society (e.g. [27,28]) was the second-highest journal followed by the European Journal of Information Systems (e.g. [29,30]).

The rest are Discover Artificial Intelligence (2022), Journal of Industrial Information Integration, MIS Quarterly, Engineering Applications of Artificial Intelligence, Frontiers in Medicine, Frontiers in

Artificial Intelligence, The Innovation, Intelligence, Electronic Markets, IEEE Transactions on Industrial Informatics, Artificial Intelligence in Medicine, Journal of Computer Information Systems, MIS Quarterly Executive, Computers in Human behaviour, Science China Information Sciences, International Journal of Information Management, Journal of Science Education and Technology, Interactive Learning Environments, Information Systems Management, Technological Forecasting and Social Change and Critical Care. The review found that few publications have been recorded and are accessible in the basket of eight (8) information systems. A possible reason is that, since AI research moves quickly, some good journals often have longer review times, which might delay the latest advancement in these journals. As individuals and organisations continue to adopt it, more publications will be recorded in the top information systems journals and in real-time for further reviews and insights.

4.2. Dominant themes

As stated earlier, the dominant themes are classified based on the typologies of [25] to reflect the current development of AI. The breakdown of articles classified under different themes in AI research is as follows. Themes under technological issues were (36.5%), contextual issues (25.9%), domain and application (23.5%) and AI conceptualization represented 12.9%. Considering that AI is underpinned by technical elements like machine learning, deep learning, NLP, speech recognition, fuzzy logic and expert systems, it comes as no surprise that technological issues theme comprise the majority of the literature. The prevalence of the technological issues theme also highlights the relatively lower focus on other themes, including AI contextual issues, conceptualisation, and application domains (see Fig. 2).

Table 2
Artificial Intelligence Framework Taxonomy.

Themes	Sub-themes
Technological issues	Data quality and quantity, robustness, bias, model interpretability, scalability, ethical AI design, explainable AI (XAI), AI governance and regulation
Contextual knowledge co-creation issues	Ethical considerations, privacy concerns, bias and fairness, job displacement, social acceptance, cost, regulatory and legal challenges
Conceptualisation	AI safety and control, AI predictability, ethical AI frameworks
Domains and applications	e-health (disease diagnosis), fraud detection, Chatbots – NLP, robotics, computer vision, game design, personalized learning, cybersecurity, smart grids

Table 3
Strings considered in the search process.

Scientific database	Search string
Scopus	TITLE-ABS-KEY (“Artificial intelligence” OR “Machine learning” AND “Information systems” OR “Information technology” “Innovation” “AI Domains” “AI +Technological Issues” “AI Applications” “AI +Machine”
AISeLib, Web of Science	TS =(“Artificial intelligence” OR “Machine learning” AND “Information systems” OR “Information technology” “AI” AND

“Contextual” OR “Conceptualisation” “Frameworks”

The Technical or technological issues theme focuses on infrastructure that supports data quality and availability. Issues examined under this theme include AI scalability and explainable AI, security, robustness and architecture, and interpretability. AI data quality and scalability studies (e.g. [19,31]) shed light on the models and data that support model construction and argue for improvement in defining, identifying and explaining errors in data. These studies also point out that AI constantly requires vast amounts of high-quality data to learn and make accurate predictions devoid of bias and inconsistencies leading to inaccurate results. Thus, the performance of AI needs significant improvement based on high-quality data in the form of 4 Vs (volume, velocity, veracity and variety).

Studies [26,28,32,33] relating to security, robustness and architecture of AI under technical issues identify adversarial attacks and model vulnerabilities, generalization and data privacy, and interoperability, real-time processing and model complexity as enablers for AI optimization. Furthermore, these studies identify factors such as personalization, social presence, compatibility, responsiveness and anthropomorphism for improving the sustainability of AI services. These factors are not exhaustive as contextual issues differ in implementation and adoption. Thus, AI cannot achieve true human predictive behaviour by solely relying on one paradigm but its novel development for a safe and reliable AI should be premised on a new explainable and robust AI theory.

Studies in AI service management have examined the diverse range of services delivered via AI infrastructure. Depending on most organizational AI applications and setup, these AI services fit the following cloud service models: SaaS, PaaS and IaaS [34]. Some of the AI services include customer support and chatbots, predictive analytics, recommendation systems, fraud detection, voice assistance, content generation, healthcare diagnostics and security and threat detection [35–39]. The underlying discussion amongst these studies is data protection and cloud security, standardization and regulatory frameworks and governance mechanisms.

The inherent characteristics of AI make its adoption a challenge. The recurring concern of AI is security. Even though providers of AI services continue to provide assurances and make strides in solving the challenges, the issue of security is a major hindrance amongst adopters. The

Fig 1.Review Selection Process.

Fig 2.Classification of Themes.

three areas of security concerns include data security, model security and network security. Data security emphasises the foundation of trust in AI development. Researchers in this space focus on data classification and minimization, anonymization and masking of data and developing threat detection and response mechanisms [26,38,40,41]. At the same time, the model security can be referred to as a system for guarding the gatekeepers. Research on model security, which is more progressive has focused on explainability, resilience, interpretability and continuous monitoring and retraining [10,34,35,42]. Finally, are the walls between the systems – referred to as network security. Development in network security prioritises infrastructure or API security, threat intelligence and intrusion detection systems [37,40,43,44]. Table 4 shows some major security issues of AI and measures geared towards addressing them.

The literature acknowledges that securing AI is an ongoing process that requires a multidisciplinary approach and collaboration amongst data scientists, AI researchers and legal teams.

Finally, the prevalence of technological issues is attributed to the fact that deep learning and massive amounts of data currently dominate the field. However, the overreliance on technological progress without investigating the contextual or theoretical foundations of intelligence may hinder the advancement of truly building AI responsible [30] for adapting to its environment and beyond specific issues. To address the overreliance on technological issues, AI research must promote a balanced approach of diverse AI methodologies, mitigate biases in large datasets and democratize AI research to foster open-source initiatives.

The contextual knowledge co-creation issues theme encompasses research concerning the socio-technical aspects of AI, including acceptance, adoption, integration, cost and ethics. Essentially, it involves collaboratively building knowledge specific to a particular context for a comprehensive understanding of an issue or situation. In contrast, the majority of studies in the contextual issues theme predominantly focus on the adoption and acceptance of AI in various domains, specifically examining factors influencing its acceptability in work activities. These adoption-focused studies (e.g., [11,31,43,47]) primarily explore organisational adoption of AI, with less emphasis on individual and country-level adoption.

Concerning the acceptance and adoption of AI, the review identified psychological anthropomorphic AI devices, perceived value, technology readiness, management support, compatibility and AI generalisation [11,47–49]. This shows a high degree of organizational acceptance and adoption of AI. Hindrances to AI adoption include privacy concerns, lack of education and training, transparency, trust, and cultural and social factors in some developing economies [10]. Addressing these contextual issues is crucial for fostering widespread acceptance of AI technologies and ensuring that they are developed and deployed in ways that benefit society while respecting ethical and legal considerations.

The issue of privacy and trust was well espoused in the literature [37, 46]. The primary focus of these studies revolves around addressing trust-related challenges that could emerge due to the adoption of AI. A proposed legislation, as advocated by Floridi [45] serves as protective measures for AI users and aims to mitigate fears and reduce security concerns associated with cloud innovation.

The debate about the bias and fairness of AI has received less attention, however, concerns have been raised about regulatory frameworks and legal challenges and how to ensure social acceptance [1,18, 50]. Thus, the non-existence of universal laws governing AI in education, health and organisations has been the main challenge of AI adoption (Mikalef et al. 2020; [28]). Given the intelligence level and storage of data in the cloud, more studies have called for measures to ensure confidentiality and transparency [33] of AI, especially chatbots and ChatGPT. In terms of cost, some papers debate that it is a challenge to maintain AI in the long run and suggest improved AI sustainability [32]. However, AI has lowered technology adoption costs and proved to be cost-effective for startups [38]. In short, the main barriers to AI adoption identified in the papers include cost, data security and privacy concerns, trust issues, perceived risks and job displacement.

The third examined theme is artificial intelligence conceptualisation. Articles [6,17,41,49,50] in this domain seek to provide a sub-structure and a starting point for AI. Given the nascent field of AI, it is important to explain basic concepts and processes to promote the field. This theme views AI research from three main lenses – operational understanding, technical understanding, and theoretical understanding. The operational articles seek to provide an understanding of how individuals comprehend the practical aspects of AI and how it can be deployed in various domains. Thus, studies with an operational understanding try to exemplify what AI applications can do and how they generally work [51–53]. The technical understanding of AI encompasses a deeper knowledge of the technical underpinnings of AI and how the models, algorithms and data process techniques are used in AI-based technologies. Understanding the technical issues involves hands-on

Table 4
Security Issues in Artificial Intelligence .

Security issue	Measures addressing issue	Potential implications for IS future research	Literature sources
Data security	Data encryption – encrypting sensitive data to prevent unauthorized access	Systems vulnerability to adversarial attacks developed to manipulate output or distort sensitive information. Decision-making processes may be complex to understand. However, AI can aid the advancement of security solutions in IS, e.g. threat prediction	[26,38,40, 41,45,46]
	Access control – implement strict access controls to only authorized users who can access and modify data		
	Data privacy – strict compliance with data privacy regulations to protect AI applications and user data		

Model security	Model testing and protection – thorough testing of AI models to identify vulnerabilities to ensure reliability and safety.	There is an enhanced system of trust and reliability. However, the prevalence of AI in IS creates new attacks for malicious players i.e. potential for unintended consequences and misuse	[10,34,35, 42]
Network security	Firewalls and intrusion detection and prevention – the use of firewalls and deploying of intrusion detection to monitor and block malicious activities.	Implication for enhanced data protection (i.e. secured networks) and improved research reliability and accuracy. However, excessive network security measures may hinder information sharing and collaboration. There is a need to strike a balance between security and openness to improve innovation and collaboration in IS.	[37,40,43, 44]

Table 5
Overview of research methods .

Research method	Approach	Percent
Qualitative	Survey and interviews	14.2%
	Case studies and interviews	
Quantitative	Experiment	49.4%
	Statistical analysis	
	Experiment and case study	
	Survey and experiment	
	Design science	
Conceptual	No methodology	24.7%
Other		11.7%

e.g. Design science

experience in programming AI systems [54,55]. Finally, is the theoretical understanding of AI which involves cognitive science, mathematical principles and the philosophical foundation of AI. The theoretical sub-theme provides deep questions about what intelligence is and how AI applications can replicate it [5,43,56,57].

The last theme is the application domain in which AI has been applied. The publications in this theme examine the application of AI in scientific research and medicine [35,41], education [58], e-governance [20], legal [45], industry [17], customer service [38,47,59], advisory systems [42], sustainable entrepreneurship and SDG [10,60] and computer network security [40]. The studies identified recognize the positive impact that AI has brought to these domains. In this regard, these studies promote the development, application and use of AI-based systems to support business processes in these areas. Scholars in these domains assert that the existing advantages derived from AI could be further enhanced through the development of new AI applications. Researchers are encouraged to contribute more to the less dominant areas such as transportation, green IT, agriculture, climate modelling, cybersecurity, energy and utilities, compliance and retail and e-commerce.

4.3. Overview of research methods

As shown in Table 5, different research methods have been used to study AI such as surveys, case studies, experiments, and statistical analysis. Given that AI is now gaining research momentum in information systems, most of the studies are conceptual (24.7%) and editorial in review (e.g. [1,30,50]) with less established research methodologies. For example, Dietzmann et al. [61] proposed a conceptual design cycle and iterative AI to understand possible synergies for financial services. Few of the studies (14.2%) used qualitative research (e.g. [51,53]) while majority of the studies used simulation and experimental frameworks (e.g. [33,39,42,62]). 34% of the studies either used quantitative such as experiments, surveys and statistical analysis (e.g. [32,58]) or mixed research methodologies such as surveys and interviews, and case studies (e.g. [36]). Experimental research was the most dominant research method used to conduct AI research. Thus, the experimented studies provided improved and quality data on the use of AI in context and environmental analysis. Other studies (11.7%) did not include any methodologies, possibly because of the infant nature of the AI domain. Interestingly, previous literature reviews [1,7] did not include research methodologies, hence, this study provides a good starting point for research into AI. Consequently, calls are made for the use of well-established research methods in future to advance AI research. Interestingly, several research used multiple techniques as a methodological approach in AI research for gathering data. Surveys were the underlying research methodology for both qualitative and quantitative methods (see Table 5). The most popular research method for AI research is experimentation. As AI continues to advance in IS field, experiments and surveys play a crucial role in testing and validating hypotheses, understanding AI behaviour in context, benchmarking and comparing algorithms and discovering unforeseen biases and challenges.

4.4. Level of analysis

As shown in Fig 3, the majority of the studies focused on how organisations (32.9%) implement and use AI as compared to the individual level (22.4%) of analysis where investigations delved into user experiences. The overreliance on organizational research underscores the

Fig. 3.Literature Classification of Level of Analysis.

Table 6

Research frameworks.

Research framework No Percent Examples of
some papers
advancing AI-
based theory
Theoretical
advances of AI
(boundaries)
No theory 21 24.7%
Technology
acceptance model
10 11.8% Pillai, R. and
Sivathanu, B.
[68]
Autonomy
Characteristics
Social context
Generative AI
AI and human
integration
Manual generation
of reasoning
AI affordances
driving the pace
Understanding
anthropomorphism
Fuzzy logic 6 7.1% Spengler, T.,
Volkmer, T., &
Herzog, S. [4]
Self-determination
theory (SDT)
4 4.7% Jimenez-
Barreto et al.
[69]
Critical theory 5 5.9% Grover, V., &
Lyytinen, K.
[65].
Anthropomorphism
theory
5 5.9% Han, M. C. [70]
Blut, Wang,
Wünderlich &
Brock [2]
Knowledge-based
View
9 10.6% Cooper,
Pereira,
Vrontis, & Liu,
[66].
Learnability
Characteristics
Human-driven data

analysis
Structured data
Adversarial learning
Extending the
resources and
knowledge
Insights from new
contexts
Interpretability
Performance
Social presence theory 6 7.1% Jiang et al.
[56]
Unified theory of
acceptance and use
of technology
(UTAUT
5 5.9% Moriuchi [71]
Social response theory
(SRT)
4 4.7% Adam, Wessel
& Benlian.
[72]
Social cognitive theory
(SCT)
4 4.7% Henschel, R.
Hortensius, E.
[67]
Uses and gratification
theory (UGT)
1 1.2% Chang et al.
[73]

continuing recognition of the significance of IS within organisations. Furthermore, a substantial number of the papers fell under the general studies category (30.6%) due to the complexity and technicality of the area. Thus, these studies do not fall under organizational (macro level), country (meso level- 14.1%) or individual level (micro level) analysis. As such they provide a general insight into AI research. More studies are needed on AI research at the individual, country, and global levels to bridge the research gaps [20,35,49] in understanding the various classification levels within AI.

4.4.1.Impact of AI on level of analysis

Technological developments have played a pivotal role in shaping organisational processes and procedures since the advent of email communication. In recent times, the advancement of these innovations has promoted digital transformation initiatives that facilitate the exchange of information and prominence in business strategy. However, the integration of data science and machine learning, and artificial intelligence (AIML) has advanced the efficiency of business operations [3]. The current trend of AI has had a positive impact on business performance and management, distribution channels and pricing strategies and not only the improvement of information quality [3]. However, the over-reliance on AI-based technologies may result in added pressure on employees and a risk of disconnection from potential customers or users. For example, Ofosu-Ampong & Acheampong [63] found that while organisations promoted technology-based solutions during the advent of covid-19 pandemic, users were reluctant to change their mode of preferred operation and engagement. In this regard, more research is needed to investigate new business models with AI and how it can strengthen firm-customer relationships to prevent wastage. Thus, AI should promote opportunities that strategically manage user or organizational retention [64], moving beyond the limitation of discontinuance of service or relationship. However, as AI continues to improve in speed, efficiency and availability, the competitive advantage for individuals and organisations will be centred on sustainability. Ultimately, organisations that invest hugely in the dynamism of human-technology interaction may control the market, irrespective of the level of analysis.

4.5. Research framework

The categorization under this section was based on the theories and models found in the publications reviewed. Due to the infant area of AI, most of the studies did not use theories or frameworks (e.g. [57]) but tried to propose a supporting theory. However, technology acceptance (TAM) was predominantly used to examine the acceptance, performance and human integration with AI. Other frameworks underlying AI include fuzzy logic, self-determination theory, critical theory and anthropomorphism theory.

Interestingly, two themes (i.e. autonomy and learnability) emerged from the conceptualization of the research framework on AI. To begin with the autonomy of AI, these theoretical themes comprise publications (e.g. [2,4,65]) that seek to provide the autonomic behaviour of AI based on the human and social experiences. The scope dimensions of these frameworks for studying AI refer to the autonomous characteristics of continuous growth. The primary articles on autonomous behaviour aim to provide an understanding of the foundational components of AI. Understanding the building blocks of AI facilitates its contextualisation

Table 7
Contemporary AI and Earlier AI in Organisational Learning.

Organizational learning	Earlier Iterations of AI	Contemporary AI
Decision-making capabilities	Primarily respond to a programmed request	Independent decision-making
Level of learning	Primarily static and rule-based	Adaptation and continuous learning
Results-orientated	Self-determined goals	Rigid predefined goals
Human interaction	Significant human interaction	Reduce human interaction

Fig 4. Artificial Intelligence Framework.

and fosters growth. The underlying contextual meaning of these frameworks is that AI permeates almost all aspects of human and social experiences.

The second half of the models and theories underlying AI research comprises of knowledge-based view, social present theory, unified theory of acceptance and use of technology (UTAUT), social response theory (SRT), social cognitive theory (SCT) and uses and gratification theory (UGT). The emerging theme from this research framework is the learnability of AI. This theme comprises articles (e.g. [56,66,67] that show the predicative and learnability characteristics of AI. Given the infant area of AI, it has become important to explore human-driven data analysis, the structure of data, adversarial learning and the interoperability of AI. The next section provides more insight into the theoretical dimension of autonomy and learnability of AI. Table 6 shows the research frameworks and the corresponding theoretical advancement of AI.

4.6. Theoretical advances of AI

From the literature review, two main frontiers of AI were identified to advance the theoretical understanding of intelligence and to explore AI's future capabilities – i.e. AI is emerging. Contemporary AI, which continually drives the boundaries of computational capabilities to tackle increasingly intricate decision-making challenges, distinguishes itself from earlier iterations in two primary aspects that significantly affect organizational learning in dealing with AI's potential: autonomy and learnability. The discussion follows as summarized in Table 6.

4.6.1. Autonomy

Recent advancements in AI exhibit a propensity to operate autonomously, making decisions and taking actions in the real world that yield tangible consequences, frequently occurring not only without human intervention but also beyond human awareness [57]. For example, Chatbots powered by AI can engage in conversations with users, answer questions, and provide assistance without human operators – using natural language processing to understand and respond to user inquiries [24].

4.6.2. Learnability

Learnability is the machine's ability to refine or improve automatically through data [74]. From its inception, the foundational principle in AI has been autonomously enhancing performance through data and experience, and we now have a firm grasp of the core principles in both supervised and unsupervised learning [74]. However, recent advancements in fields like deep learning and reinforcement learning have become achievable primarily because of the vast availability of big data [75]. The enhanced learning capabilities of AI now allow it to excel in more intricate decision-making scenarios, such as tasks involving audio, speech, object recognition, and natural language processing [75,76]. For example, Language models like GPT-3 can learn from a massive amount of text data and adapt to understand and generate human-like text – to become better at tasks like language translation, text summarization, and even creative writing.

4.6.3. So how does contemporary AI research differ from earlier iterations?

In the spheres of learnability, contemporary AI focuses on enabling dynamic decision-making with emphasis on algorithms that endlessly improve performance with large datasets. Technologies on display include deep learning, neural networks and reinforcement learning that focus on adapting to environmental changes [77,78]. Earlier AI mostly relied on pre-programmed decision trees and rules, thereby limiting its learning potential. Its limitations also included specific tasks and the inability to adjust to new data or unforeseen prompts [79].

Autonomous AI in contemporary AI research in IS focuses on
Table 8
Future Direction of AI Research.
AI Theme Challenges with AI Research gaps Future research direction
Technological issues There are growing concerns about the handling of sensitive data in AI systems due to potential misuse and breaches.
The prevalence of AI systems is raising more concerns with regulatory frameworks and ethics.
•Most of the AI research focuses on the impact of AI on performance
•Most of the current AI research focuses on security issues
•Need for regulatory framework and ethical considerations to ensure responsible deployment and adoption
•Future research is required to explore the mediating role of AI's impact on performance e.g. academic or firm performance.
•Interestingly, research is needed on how human-AI integration influences human-to-human interaction. To build trust, more research is needed on robust security measures and transparency
Contextual knowledge co-creation issues
Contextually, AI applications can learn unintended and unforeseen behaviours from unstructured and complex data. This can lead to harmful or unintended consequences in societies.
•Limited AI research related to context and generalization.
•Many studies have been investigated at the organizational level rather than country level
•Contextual and continuous monitoring to refine the AI model is key to its sustainability
•Future research needs to investigate the sustainability of AI per context and compare results e.g. developing vs developed country use of AI.
•Further, research should verify AI value and user retention via context
•How do we combine different perspectives to build trust and collaboration with AI
Conceptualisation AI models can act as "black boxes" making it difficult to understand its conceptualization and decision-making processes.
•Few AI research focuses on the theoretical dimensions of AI due to its infancy
•Limited studies on LLMs performance compared to traditional chatbots in delivery outcomes
•Future research is required to test new research frameworks and also explore integrative approaches to develop new ones to enhance our

understanding of AI’s future directions
Domains AI applications and data required have the potential to grow. However, there are issues with data storage, processing and analysis of infrastructure needed to manage the growth of AI (i.e. the domain of AI is showing an ever-increasing flow)
•Most of the AI research focuses on the various domains of AI and how it can be applied
•Also, AI’s ability to predict future events and outcomes is still an open research agenda
•Future research is required to venture into new unexplored domains like sustainability and green IT.
•Future research is required to advance our knowledge of how domain models differ from AI-tailored solutions
General recommendation for businesses •The potential of AI workforce disruption and ethical issues requires an improved learning outcome, increased efficiency and research advancement
•Discussing the future of business operation’s best practices, challenges and opportunities
•Embrace individual changes, analytics and learning preferences and value the role of humans
•Investing, recruiting and retaining employees who can design, and model AI-tailored solutions to business goals

designing AI applications capable of independent decision-making e.g. self-learning and goal-orientated behaviour [19,41]. While the earlier AI were predominantly reactive systems that responded to predetermined instructions [80]. Thus, human control and oversight dominated earlier AI – and could not plan and execute actions independently. Table 7 shows the impact of contemporary AI and earlier AI on organizational learning.

Overall, the gradual shift towards advanced autonomy and learnability in AI reflects the enhanced computational power and algorithms.

4.7. Towards artificial intelligence framework

From the review, an inductive framework is developed to synthesize the extant knowledge of current AI in an organisational environment. As summarized in Fig. 4, the main themes comprise technological issues, contextual issues, conceptualization and domain issues while the main outcomes of AI is the autonomy and learnability. The linkages from Fig. 4 emerged from the studies analysis and suggest that in a broader organizational or country-level development and use of AI, the four thematic issues may trigger a set of digital transformations initiated by AI which may yield an autonomous or learnability pattern of the institution. In essence, these factors or themes may shape or reconfigure the transformation process of AI; however, real-time learnability results can be directly derived from the various domains. The proposed AI framework aims to help identify research gaps for future investigations. Thus, future research in IS can explore the different interrelationships between the identified themes and outcomes in Fig. 3 to investigate critical issues affecting AI.

5. Implications and future research directions

This review has implications for academics, practitioners and policymakers in AI development and deployment.

5.1. Academics

Across the disciplines, AI is fast transforming research methodologies which has implications for data analysis, research gap identification, literature analysis and personalization of learning and research experiences. This unravels the new discovery of management information systems and innovation. However, it raises concerns about human elements in the research process, ethical data practices and research reproducibility.

5.2. Practitioners

As AI continues to automate roles and activities, there is a need to develop the workforce to adapt and acquire new skills e.g. data analytics and critical thinking to avoid job displacement and social inequalities. To build trust and ensure fairness and accountability, responsible AI practices need to be implemented. To advance the implementation of responsible AI and set standards for data privacy and safety, Policymakers need to develop regulatory frameworks. These regulations can address the current challenges of ethical, legal and social posed by AI.

5.3. Limitation

This review only considered peer-reviewed articles, hence falls short of some literature and studies. Also, since the focus was on a selected number of IS-related journals and articles, some studies from non-IS outlets and IS conferences were excluded. Table 8 shows emerging themes, challenges, research gaps, and future research directions of AI in advancing IS research.

6. Conclusion

This article examined artificial intelligence literature through the analysis of 85 articles to determine the current issues and stock of knowledge in AI literature, research methodology, level of analysis and conceptual approaches to identify research gaps for future investigations. The main contribution of this study is the themes' classifications into i) technological issues, ii) contextual issues, iii) domain and application and iv) conceptualisation. The Technical or technological issues theme focuses on infrastructure that supports data quality and availability while the contextual issues theme encompasses research concerning the socio-technical aspects of AI, including acceptance, adoption, integration, cost and ethics. From the analysis, two main frontiers of AI were identified to advance the theoretical understanding of intelligence and also to explore AI's future capabilities – i.e. autonomy and learnability. The study concludes with an AI framework and research gaps to support the growth of AI in business processes and procedures. As shown in Appendix A, the takeaway from the article summary is the research interest in security and ethics, business model focus, digital transformation and innovation (reshaping value propositions), platform growth, sourcing AI evolution, continued themes with new framing (IT/AI strategy and alignment), AI autonomy and knowledge creation. Finally, the researcher hopes this study will spark further exploration in AI research.

CRediT authorship contribution statement

Kingsley Ofosu-Ampong: Conceptualization, Methodology, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence
the work reported in this paper.
Data availability
No data was used for the research described in the article.
Appendix A.Articles by journal publication and research issues
Journal Some names of
authors
Research issues
IEEE Transactions on Industrial Informatics [17] AI to explainable AI
Security and ethics
Computers in Human behaviour (x2) [59] AI-enabled customer experiences Trust-commitment theory and service quality model
International Journal of Intelligent Information
Systems
[42] AI Chatbot Advisory System Platform growth
MIS Quarterly Executive [5] AI and social dysfunction Authority reporting
Journal of Data and Information Quality [19] Data quality and explainable AI
(continued on next page)

(continued)
Journal Some names of
authors
Research issues
Qualitative Research in Financial Markets [51] Robo-advisors Automated online advisory platforms
Journal of the Academy of Marketing Science [2] AI and anthropomorphism Service provision; Physical robots
Decision Support Systems (x4) [81] Perceived customer autonomy Explainable AI and customer emotions
Business Horizons (x2) [3] A framework for diagnosing value destruction potential
[58] AI and governance change structure Tailored internal architecture UTAUT, Issues of autonomy
European Journal of Information Systems (x3) [29] AI and self-disclosure of personal information Privacy and participation in ridesharing
Artificial Intelligence in Medicine [35] Explainability of artificial intelligence in medicine
Journal of Science Education and Technology [52] Generative artificial intelligence New business model
Journal of Business Research (x2) [66] Resource and knowledge-based view
International Journal of Information Management
(x3)
[47] Intelligent service robots AI trust and culture
Telematics and Informatics (x3) [48] Technology acceptance theories AI-based intelligent products
Journal of Information Technology(x2) [65] Innovativeness in the digital age
Technological Forecasting and Social Change (x2) [60] AI-based technologies Sustainable entrepreneurship
Journal of Strategic Information Systems [82] ML value creation; Knowledge creation Task augmentation Human-in-the-loop work configuration
Digital/human work configuration
Journal of Internet Commerce (x2) [70] AI anthropomorphism Purchase decision in chatbot commerce
Journal of Business Logistics [53] Technology acceptance model, artificial intelligence and trust issues
Journal of Intelligent & Fuzzy Systems [40] Computer network security and artificial intelligence
Engineering Applications of Artificial Intelligence [43] Artificial intelligence Metaverse development
Frontiers in artificial intelligence (x2) [83] AI trustworthiness, explainability, and ethics Implicit anthropocentric anthropomorphistic concepts
Communication Studies [46] Privacy and machine values
Neural Networks [75] Deep learning, reinforcement learning World models
European Journal of Information Systems (x4) [30] Responsible AI Dark side of AI
Nature Machine Intelligence [26] Skills for physical AI
Academy of Management Review [57] Conjoined agency
Information and Knowledge Management. [22] AI trust, innovativeness Psychological needs

Journal of Computer Information Systems (x2) [32] AI, systemic factors and chatbot sustainability
Journal of Retailing and Consumer Services [21] Autonomous vehicles and demographic variables
Computers in Human behaviour (x4) [31,36] AI device human-like; Empathy and perceived psychological anthropomorphism AI development with
social chatbots
International Journal of Contemporary Hospitality
Management
[68] AI-based chatbots; Hospitality and tourism integration with AI
AI & Society (x5) [27]. Psychoanalysing of AI – Replika
[28] AI policy, ethics, and regulation
Learning and Instruction [62] Adaptive feedback AI neural networks Pre-service teachers' diagnostic reasoning
Critical Care [54] AI and scientific writing
Information Systems Management (x2) [20] AI and governance in businesses.
Expert Systems with Applications [78] Explainable AI Intrusion detection in IoT networks Deep learning-based approach
Journal of Service Management [38] AI +feeling Customer experience with chatbot
Electronics (x2) [39] IoMT +deep CNN; Pandemic diseases AI-based intelligent support system
Electronic Market [33] Trustworthy AI
AI and Ethics [55] Sustainable AI
Intelligence [49] Intelligence level of AI strategy and alignment
Nature Machine Intelligence (x11) [6] A new generation of AI
The Innovation (x2) [41] AI and scientific research
International Journal of Electrical, Electronics and
Computers
[34] AI modelling in ERP Cloud-Based System
Science China Information Sciences [84] 3rd generation AI

References

[1] N. Berente, B. Gu, J. Recker, R. Santhanam, Managing artificial intelligence, MIS Q. 45 (3) (2021).
[2] M. Blut, C. Wang, N.V. Wünderlich, C. Brock, Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI, J. Acad. Market. Sci. 49 (2021) 632–658.
[3] A.I. Canhoto, F. Clear, Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential, Bus Horiz 63 (2) (2020) 183–193.
[4] T. Spengler, T. Volkmer, S. Herzog, Artificial intelligence and Fuzzy logic in modern human resource management, Working Paper Ser. (2023).
[5] H. Benbya, T.H. Davenport, S. Pachidi, Artificial intelligence in organizations: current state and future opportunities, MIS Q. Execut. 19 (4) (2020).
[6] F. Wu, C. Lu, M. Zhu, H. Chen, J. Zhu, K. Yu, Y. Pan, Towards a new generation of artificial intelligence in China, Nature Mach. Intell. 2 (6) (2020) 312–316.
[7] C. Zhang, Y. Lu, Study on artificial intelligence: the state of the art and future prospects, J. Ind. Inf. Integrat. 23 (2021) 100224.
[8] J. Chappell, Z.P. Demery, V. Arriola-Rios, A. Sloman, How to build an information gathering and processing system: lessons from naturally and artificially intelligent systems, Behav. Processes 89 (2) (2012) 179–186.
[9] Y. Bathaee, The artificial intelligence black box and the failure of intent and causation, Harv. JL Tech. 31 (2017) 889.
[10]A. Di Vaio, R. Palladino, R. Hassan, O. Escobar, Artificial intelligence and business models in the sustainable development goals perspective: a systematic literature review, J. Bus. Res. 121 (2020) 283–314.
[11]S. Kelly, S.A. Kaye, O. Oviedo-Trespalacios, What factors contribute to acceptance of artificial intelligence? A systematic review, Telematic. Inf. (2022) 101925.
[12]K. Ofosu-Ampong, Gender differences in perception of artificial intelligence-based tools, J. Digit. Art Humanit. 4 (2) (2023) 52–56.
[13]C. Collins, D. Dennehy, K. Conboy, P. Mikalef, Artificial intelligence in information systems research: a systematic literature review and research agenda, Int. J. Inf. Manage. 60 (2021) 102383.

[14]P. McCorduck, Machines Who think: A personal Inquiry into the History and
Prospects of Artificial Intelligence (2nd ed.). A. K. Peters, 2004.
[15]A.M. Turing, I.—Computing Machinery and Intelligence, Mind, LIX, 1950,
pp. 433–460.
[16]S. Russel, P Norvig, Artificial Intelligence: A Modern Approach (4th ed.), N.J.:
Prentice Hall, 2020.
[17]I. Ahmed, G. Jeon, F. Piccialli, From artificial intelligence to explainable artificial
intelligence in industry 4.0: a survey on what, how, and where, IEEE Trans. Ind.
Inf. 18 (8) (2022) 5031–5042.
[18]S. Feuerriegel, Y.R. Shrestha, G. von Krogh, C. Zhang, Bringing artificial
intelligence to business management, Nature Mach. Intell. 4 (7) (2022) 611–613.
[19]L. Bertossi, F. Geerts, Data quality and explainable AI, J. Data Inf. Qual. (JDIQ) 12
(2) (2020) 1–9.

[20]J. Schneider, R. Abraham, C. Meske, J. Vom Brocke, Artificial intelligence
governance for businesses, Inf. Syst. Manag. 40 (3) (2023) 229–249.
[21]J. Park, E. Hong, H.T. Le, Adopting autonomous vehicles: the moderating effects of
demographic variables, J. Retail. Consum. Serv. 63 (2021) 102687.
[22]K. Ofosu-Ampong, B. Acheampong, M.O. Kevor, F. Amankwah-Sarfo, Acceptance
of artificial intelligence (ChatGPT) in education: trust, innovativeness and
psychological need of students, Inf. Knowl. Manag. 13 (4) (2023) 37–47.
[23]A.F. Borges, F.J. Laurindo, M.M. Spínola, R.F. Gonçalves, C.A. Mattos, The strategic
use of artificial intelligence in the digital era: systematic literature review and
future research directions, Int. J. Inf. Manage. 57 (2021) 102225.
[24]T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, S. Bisen, Implementation of a Chatbot
System using AI and NLP, Int. J. Innovat. Res. Comput. Sci. Technol. (IJIRCST) 6
(3) (2018).
[25]H. Yang, M. Tate, A descriptive literature review and classification of cloud
computing research, Commun. Assoc. Inf. Syst. 31 (1) (2012) 2.
[26]A. Miriyev, M. Kova
,p
c, Skills for physical artificial intelligence, Nature Machine
Intelligence 2 (11) (2020) 658–660.
[27]L.M. Possati, Psychoanalyzing Artificial Intelligence: The case of Replika, 38, AI &
Society, 2023, pp. 1725–1738.
[28]H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang, L. Floridi, The Chinese
approach to artificial intelligence: an analysis of policy, ethics, and regulation AI
Soc. 36 (2021) 59–77.
[29]X. Cheng, L. Su, X. Luo, J. Benitez, S. Cai, The good, the bad, and the ugly: impact
of analytics and artificial intelligence-enabled personal information collection on
privacy and participation in ridesharing, Eur. J. Inf. Syst. 31 (3) (2022) 339–363.
[30]P. Mikalef, K. Conboy, J.E. Lundstr
0€
om, A. Popovi
,p
c, Thinking responsibly about
responsible AI and 'the dark side'of AI, Eur. J. Inf. Syst. 31 (3) (2022) 257–268.
[31]C. Pelau, D.C. Dabija, I. Ene, What makes an AI device human-like? The role of
interaction quality, empathy and perceived psychological anthropomorphic
characteristics in the acceptance of artificial intelligence in the service industry,
Comput. Human Behav. 122 (2021) 106855.
[32]A. Park, S.B. Lee, Examining AI and systemic factors for improved chatbot
sustainability, J. Comput. Inf. Syst. (2023) 1–15.
[33]S. Thiebes, S. Lins, A. Sunyaev, Trustworthy artificial intelligence, Electron.

Market. 31 (2021) 447–464.
[34]N. Yathiraju, Investigating the use of an artificial intelligence model in an ERP cloud-based system, Int. J. Electric. Electron. Comput. 7 (2) (2022) 1–26.
[35]C. Combi, B. Amico, R. Bellazzi, A. Holzinger, J.H. Moore, M. Zitnik, J.H. Holmes, A manifesto on explainability for artificial intelligence in medicine, Artif. Intell. Med. 133 (2022) 102423.
[36]I. Pentina, T. Hancock, T. Xie, Exploring relationship development with social chatbots: a mixed-method study of replika, Comput. Human Behav. 140 (2023) 107600.
[37]S.S. Kute, A.K. Tyagi, S.U. Aswathy, Security, privacy and trust issues in internet of things and machine learning based e-healthcare, Intell. Interact. Multimed. Syst. e-Healthcare Appl. (2022) 291–317.
[38]K. Sidaoui, M. Jaakkola, J. Burton, AI feel you: customer experience assessment via chatbot interviews, J. Serv. Manag. 31 (4) (2020) 745–766.
[39]S. Thandapani, M.I. Mahaboob, C. Iwendi, D. Selvaraj, A. Dumka, M. Rashid, S. Mohan, IoMT with deep CNN: aI-based intelligent support system for pandemic diseases, Electronics (Basel) 12 (2) (2023) 424.
[40]T. Hua, L. Li, Computer network security technology based on artificial intelligence, J. Intell. Fuzzy Syst. 37 (5) (2019) 6021–6028.
[41]Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, J. Zhang, Artificial intelligence: a powerful paradigm for scientific research, Innovation 2 (4) (2021).
[42]C.U. Betrand, O.U. Ekwealor, C.J. Onyema, Artificial intelligence Chatbot advisory system, Int. J. Intell. Inf. Syst. 12 (1) (2023) 1.
[43]T. Huynh-The, Q.V. Pham, X.Q. Pham, T.T. Nguyen, Z. Han, D.S. Kim, Artificial intelligence for the metaverse: a survey, Eng. Appl. Artif. Intell. 117 (2023) 105581.
[44]J.H. Li, Cyber security meets artificial intelligence: a survey, Front. Inf. Technol. Electron. Eng. 19 (12) (2018) 1462–1474.
[45]L. Floridi, The European legislation on AI: a brief analysis of its philosophical approach. The 2021 Yearbook of the Digital Ethics Lab, Springer International Publishing, Cham, 2022, pp. 1–8.
[46]J. Lobera, C.J. Fern
0
andez Rodríguez, C. Torres-Albero, Privacy, values and machines: predicting opposition to artificial intelligence, Commun. Stud. 71 (3) (2020) 448–465.
[47]O.H. Chi, C.G. Chi, D. Gursoy, R. Nunkoo, Customers' acceptance of artificially intelligent service robots: the influence of trust and culture, Int. J. Inf. Manage. 70 (2023) 102623.
[48]K. Sohn, O. Kwon, Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products, Telematic. Inf. 47 (2020) 101324.
[49]H.L. van der Maas, L. Snoek, C.E. Stevenson, How much intelligence is there in artificial intelligence? A 2020 update, Intelligence 87 (2021) 101548.
[50]G. Briganti, O. Le Moine, Artificial intelligence in medicine: today and tomorrow, Front. Med. (Lausanne) 7 (2020) 27.
[51]A. Bhatia, A. Chandani, R. Atiq, M. Mehta, R. Divekar, Artificial intelligence in financial services: a qualitative research to discover robo-advisory services, Qual. Res. Financ. Market. 13 (5) (2021) 632–654.
[52]G. Cooper, Examining science education in chatgpt: an exploratory study of generative artificial intelligence, J. Sci. Educ. Technol. 32 (3) (2023) 444–452.
[53]A. Hasija, T.L. Esper, In artificial intelligence (AI) we trust: a qualitative investigation of AI technology acceptance, J. Bus. Logistic. 43 (3) (2022) 388–412.
[54]M. Salvagno, F.S. Taccone, A.G. Gerli, Can artificial intelligence help for scientific writing? Critical Care 27 (1) (2023) 1–5.
[55]A. van Wynsberghe, Sustainable AI: AI for sustainability and the sustainability of AI, AI Ethics 1 (3) (2021) 213–218.
[56]Y. Jiang, X. Li, H. Luo, S. Yin, O. Kaynak, Quo vadis artificial intelligence? Discover Artific. Intell. 2 (1) (2022) 4.

[57]A. Murray, J.E.N. Rhymer, D.G. Sirmon, Humans and technology: forms of conjoined agency in organizations, Acad. Manage. Rev. 46 (3) (2021) 552–571.
[58]S. Chatterjee, K.K. Bhattacharjee, Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling, Educ. Inf. Technol. 25 (2020) 3443–3463.
[59]N. Ameen, A. Tarhini, A. Reppel, A. Anand, Customer experiences in the age of artificial intelligence, Comput. Human Behav. 114 (2021) 106548.
[60]B.B. Gupta, A. Gaurav, P.K. Panigrahi, V. Arya, Analysis of artificial intelligence-based technologies and approaches on sustainable entrepreneurship, Technol. Forecast. Soc. Change. 186 (2023) 122152.
[61]C. Dietzmann, R. Heines, R. Alt, The convergence of distributed ledger technology and artificial intelligence: an end-to-end reference lending process for financial services, in: Proceedings: Twenty-Eighth European Conference on Information Systems (ECIS2020), 2020. Association for Information Systems.
[62]M. Sailer, E. Bauer, R. Hofmann, J. Kiesewetter, J. Glas, I. Gurevych, F. Fischer, Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning, Learn. Instr. 83 (2023) 101620.
[63]K. Ofosu-Ampong, B. Acheampong, Adoption of contactless technologies for remote work in Ghana post-Covid-19: insights from technology-organisation-environment framework, Digital Business 2 (2) (2022) 100023.
[64]C. Singh, M.K. Dash, R. Sahu, A. Kumar, Artificial intelligence in customer retention: a bibliometric analysis and future research framework, Kybernetes (2023), https://doi.org/10.1108/K-02-2023-0245.
[65]V. Grover, K. Lyytinen, The pursuit of innovative theory in the digital age, J. Inf. Technol. 38 (1) (2023) 45–59.
[66]C. Cooper, V. Pereira, D. Vrontis, Y. Liu, Extending the resource and knowledge-based view: insights from new contexts of analysis, J. Bus. Res. 156 (2023) 113523.
[67]A. Henschel, R. Hortensius, E.S. Cross, Social cognition in the age of human–robot interaction, Trends Neurosci. 43 (6) (2020) 373–384.
[68]R. Pillai, B. Sivathanu, Adoption of AI-based chatbots for hospitality and tourism, Int. J. Contemp. Hospital. Manag. 32 (10) (2020) 3199–3226.
[69]J. Jim
0
enez-Barreto, N. Rubio, S. Molinillo, Find a flight for me, Oscar!" motivational customer experiences with chatbots, Int. J. Contemp. Hosp. Manag. 33 (11) (2021) 3860–3882.
[70]M.C. Han, The impact of anthropomorphism on consumers' purchase decision in chatbot commerce, J. Internet Commerce 20 (1) (2021) 46–65.
[71]E. Moriuchi, An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior, Psychol. Mark. 38 (1) (2021) 21–42.
[72]M. Adam, M. Wessel, A. Benlian, AI-based chatbots in customer service and their effects on user compliance, Electron. Mark. 31 (2) (2021) 427–445.
[73]Y. Chang, S. Lee, S.F. Wong, S.P. Jeong, AI-powered learning application use and gratification: an integrative model, Inform. Techno. People 35 (7) (2022) 2115–2139.
[74]J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[75]Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, J. Morimoto, Deep learning, reinforcement learning, and world models, Neural Netw. 152 (2022) 267–275.
[76]D. Hafner, J. Pasukonis, J. Ba, T. Lillicrap, Mastering diverse domains through world models, arXiv preprint (2023). arXiv:2301.04104.
[77]C.C. Aggarwal, Neural Networks and Deep Learning, 10, Springer, 2018, p. 3.
[78]B. Sharma, L. Sharma, C. Lal, S. Roy, Explainable artificial intelligence for intrusion detection in IoT networks: a deep learning based approach, Expert Syst. Appl. 238 (2024) 121751.
[79]A. Wolfe, Mind, self, society, and computer: artificial intelligence and the sociology

of mind, Am. J. Sociol. 96 (5) (1991) 1073–1096.
[80]S.D. Whitehead, D.H. Ballard, A role for anticipation in reactive systems that learn, in: Proceedings of the Sixth International Workshop on Machine Learning, Morgan Kaufmann, 1989, pp. 354–357.
[81]Y. Guo, Y. Li, D. Liu, S.X. Xu, Measuring service quality based on customer emotion: an explainable AI approach, Decis. Support. Syst. 176 (2023) 114051.
[82]A. Shollo, K. Hopf, T. Thiess, O. Müller, Shifting ML value creation mechanisms: a process model of ML value creation, J. Strategic Inf. Syst. 31 (3) (2022) 101734.
[83]J.H. Korteling, G.C. van de Boer-Visschedijk, R.A. Blankendaal, R.C. Boonekamp, A.R. Eikelboom, Human-versus artificial intelligence, Front. Artific. Intell. 4 (2021) 622364.
[84]B. Zhang, J. Zhu, H. Su, Toward the third generation artificial intelligence, Sci. China Inf. Sci. 66 (2) (2023) 121101.
Dr. Kingsley Ofosu-Ampong is a behavioural scientist at the Alliance of Bioversity In
-
ternational and International Center for Tropical Agriculture (CIAT) in Accra, Ghana. With a PhD in Information Systems and a degree in MBA and Sociology from the University of Ghana. His research interests include behaviour change, information systems, sustain-
ability, technology adoption and agronomic behaviours.
K. Ofosu-Ampong

Compared Document:

NBER WORKING PAPER SERIES THE IMPACT OF ARTIFICIAL INTELLIGENCE ON INNOVATION Iain M. Cockburn Rebecca Henderson Scott Stern Working Paper 24449 http://www.nber.org/papers/w24449 NATIONAL BUREAU OF ECONOMIC RESEARCH 1050 Massachusetts Avenue Cambridge, MA 02138 March 2018 The authors would like to thank the organizers and participants at the first NBER conference on the Economics *of* Artificial Intelligence, and in particular our discussant Matthew Mitchell for many helpful suggestions and ideas. Michael Kearney provided extraordinary research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Funding for this paper was provided by the MIT Sloan School of Management, by the HBS Division of Research and by the Questrom School of Management. At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24449.ack NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications. © 2018 by Iain M. Cockburn, Rebecca Henderson, and Scott Stern. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source. The Impact *of* Artificial Intelligence on Innovation Iain M. Cockburn, Rebecca Henderson, and Scott Stern NBER Working Paper No. 24449 March 2018 JEL No. L1 ABSTRACT Artificial intelligence may greatly increase *the* efficiency of the existing economy. But it may have an even larger impact by serving as a new general-purpose “method of invention” that can reshape the *nature* of the innovation process and the organization of R&D. We distinguish between automation-oriented applications such as robotics and the potential for recent developments in “deep learning” to serve as a general-purpose method of invention, finding strong evidence of a “shift” in the importance of application-oriented learning research since 2009. We suggest that this is likely to lead to a significant substitution away from more routinized labor-intensive research towards research that takes advantage of the interplay between passively generated *large* datasets and enhanced prediction algorithms. *At* *the* *same* time, the potential commercial rewards from mastering this mode of research are likely to usher in a period of racing, driven by powerful incentives for individual companies to acquire and control critical *large* datasets and application-specific algorithms. We suggest that policies which encourage transparency and sharing of core datasets across both public and private actors may be critical tools for stimulating research productivity and innovation-oriented competition going forward. Iain M. Cockburn School of Management Boston University 595 Commonwealth Ave Boston, MA 02215 and NBER cockburn@bu.edu Rebecca Henderson Heinz Professor of Environmental Management Harvard Business School Morgan 445 Soldiers Field Boston, MA 02163 and NBER rhenderson@hbs.edu Scott Stern MIT Sloan School of Management 100 Main Street, E62-476 Cambridge, MA 02142 and NBER sstern@mit.edu 1 I. Introduction Rapid advances *in* *the* field *of* artificial intelligence have profound implications for the economy *as* well as society at large. These innovations *have* *the* potential to directly influence both the production and the characteristics of a wide range of products and services, with important implications for productivity, employment, and competition. But, as important as these effects are likely to be, artificial intelligence also has *the* potential to change the innovation process itself, with consequences that may be equally profound, and which may, over time, come to dominate the direct effect. Consider the case of Atomwise, a startup firm which is developing novel technology for identifying potential drug candidates (and insecticides) by using neural networks to predict the bioactivity of candidate molecules. The company reports that its deep convolutional neural networks "far surpass" *the* performance of conventional "docking" algorithms. After appropriate training on vast quantities of data, the company's AtomNet product is described as *being* able to "recognize" foundational *building* blocks of organic chemistry, and is capable of generating highly accurate predictions of the outcomes of real-world physical experiments (Wallach et al., 2015). Such breakthroughs hold out the prospect of substantial improvements in the productivity of early stage drug screening. Of course, Atomwise's technology (and that of other companies leveraging *artificial* intelligence to advance drug discovery or medical diagnosis) is still at an early stage: though their initial results seem to be promising, no new drugs have actually come to market using these new approaches. But whether or not Atomwise delivers fully on its promise, its technology is representative of the ongoing attempt to develop a new innovation "playbook", one that leverages *large* datasets and learning algorithms to engage in precise prediction of biological phenomena in order to guide design effective interventions. Atomwise, for example, is now deploying this approach to the discovery and *development* of new pesticides and agents for controlling crop diseases. Atomwise's example illustrates two of the ways in which advances *in* artificial intelligence *have* *the* potential to impact innovation. First, though the origins *of* artificial intelligence are broadly *in* *the* field of computer science, and its early commercial applications have been in relatively narrow domains such as robotics, the learning algorithms that are now being developed 2 suggest that artificial intelligence may ultimately have applications across a very wide range. From the perspective of the economics of innovation (among others, Bresnahan and Trajtenberg (1995)), there is an important distinction between the problem of providing innovation incentives to develop technologies with a relatively narrow domain of application, such robots purpose- built for narrow tasks, versus technologies with a wide—advocates might say almost limitless— domain of application, as may be true of the advances in *neural* networks *and* machine learning often *referred* to as "deep learning." As such, a first question to be asked is the degree to which developments *in* artificial intelligence are not simply examples of new technologies, but rather may be the kinds of "general purpose technologies" (hereafter GPTs) that have historically been such influential drivers of long-term technological progress. Second, while some applications *of* artificial intelligence will surely constitute lower-cost or higher-quality inputs into many existing production processes (spurring *concerns* about the potential for large job displacements), others, such as deep learning, hold out the prospect of not only productivity gains across a wide variety of sectors but also changes in the very *nature* of the innovation process within those domains. As articulated famously by Griliches (1957), by enabling innovation across many applications, the "invention of a method of invention" has *the* potential to have much larger economic impact than development of any single new product. Here we argue that recent advances in machine learning and neural networks, through their ability to improve both *the* performance of end use technologies and the *nature* of the innovation process, are likely to have a particularly large impact on innovation and growth. Thus the incentives and obstacles that may shape the development and diffusion of these technologies are an important topic for economic research, and building *an* *understanding* of the conditions under which different potential innovators are able to gain access to these tools and to use them in a pro-competitive way is a central concern for policy. This essay begins to unpack the potential impact of advances *in* artificial intelligence on innovation, and to identify the role that policy and institutions might play in providing effective incentives for innovation, diffusion, and competition in this area. We begin in Section II by highlighting the distinctive economics of research tools, of which deep learning applied to R&D problems is such an intriguing example. We *focus* on the interplay between the degree of generality of application of a new research tool and *the* role of research tools not simply in 3 enhancing *the* efficiency of research activity but in creating a new "playbook" for innovation itself. We then turn in Section III to briefly contrasting three key technological trajectories within AI—robotics, symbolic systems, and deep learning. We propose that these often conflated fields will likely play very different roles in *the* future of innovation and technical change. Work in symbolic systems appears to have stalled and is likely to have relatively little impact going forwards. And while developments in robotics *have* *the* potential to further displace human labor in the production of many goods and services, innovation in robotics technologies per se has relatively low potential to change the nature of innovation itself. By contrast, deep learning seems to be an area of research that is highly general-purpose and that has *the* potential to change the innovation process itself. We explore whether this might indeed be the case through an examination of some quantitative empirical evidence on the evolution of different areas *artificial* intelligence in terms of scientific and technical outputs of AI researchers as measured (imperfectly) by the publication of papers and patents from 1990 through 2015. In particular, we develop what we believe is the first systematic database that captures the corpus of scientific paper and patenting activity in artificial intelligence, broadly defined, and divides these outputs into those associated with robotics, symbolic systems, and deep learning. Though preliminary in nature (and inherently imperfect given that key elements of research activity *in* artificial intelligence may not be observable using these traditional innovation metrics), we find striking evidence for a rapid and meaningful shift in the application orientation of learning-oriented publications, particularly after 2009. The timing of this shift is informative, since it accords with qualitative evidence about the surprisingly strong performance of so-called "deep learning" multi-layered neural networks in a range of tasks including *computer* vision and other prediction tasks. Supplementary evidence (not reported here) *based* on the citation patterns to authors such as Geoffrey Hinton who are leading figures in deep learning suggests a striking acceleration of work in just the last few years that builds on a small number of algorithmic breakthroughs related to multi-layered neural networks. Though not a central aspect of the analysis for this paper, we further find that, whereas research on learning-oriented algorithms *has* had a slow and steady upward swing outside of the 4 United States, US researchers have had a less sustained commitment to learning-oriented research prior to 2009, and have been in a "catch up" mode ever since. Finally, we begin to explore some of the organizational, institutional and policy consequences of our analysis. We see *machine* learning as the "invention of a method of invention" whose application depends, in each case, on having access not just to the underlying algorithms but also to large, granular datasets on physical and social behavior. Developments in *neural* networks *and* machine learning thus raise the question of, even if the underlying scientific approaches (i.e., the basic multi-layered neural networks algorithms) are open, prospects for continued progress in this field—and commercial applications thereof—are likely to be significantly impacted by terms of access to complementary data. Specifically, if there are increasing returns to scale or scope in data acquisition (there is more learning to be had from the "larger" dataset), it is possible that early or aggressive entrants into a particular application area may be able to create a substantial and long-lasting competitive advantage over potential rivals merely through the control over data rather than through formal intellectual property or demand- side network effects. Strong incentives to maintain data privately has the additional potential downside that data is not being shared across researchers, thus reducing the ability of all researchers to access an even larger set of data that would arise from public aggregation. As *the* competitive advantage of incumbents is reinforced, the power of new entrants to drive technological change may be weakened. Though this is an important possibility, it is also the case that, at least so far, there seems to be a significant amount of entry and experimentation across most key application sectors. II. The Economics of New Research Tools: The Interplay between New Methods of Invention and the Generality of Innovation At least since Arrow (1962) and Nelson (1959), economists have appreciated the potential for significant underinvestment in research, particularly basic research or domains of invention with low appropriability for the inventor. Considerable insight has been gained into the conditions under which the incentives for innovation may be more or less distorted, both in terms of their overall level and in terms of the direction of that research. As we consider the potential impact of advances in AI on innovation, two ideas from this literature seem particularly 5 important—the potential for contracting problems associated with *the* development of a new broadly applicable research tool, and the potential for coordination problems arising from adoption and diffusion of a new "general purpose technology." In contrast to technological progress in relatively narrow domains, such as traditional automation and industrial robots, we argue that those areas *of* artificial intelligence evolving most rapidly—such as deep learning— are likely to raise serious challenges in both dimensions. First, consider the challenge in providing appropriate innovation incentives when an innovation has potential to drive technological and organizational change across a wide number of distinct applications. Such "general purpose technologies" (David, 1990; Bresnahan and Trajtenberg, 1995) often take *the* form of core inventions that *have* *the* potential to significantly enhance productivity or quality across a wide number of fields or sectors. David's (1990) foundational study of the electric motor showed that this invention brought about enormous technological and organizational change across sectors as diverse as manufacturing, agriculture, retail, and residential construction. Such "GPTs" are usually understood to meet three criteria that distinguish them from other innovations: they have pervasive application across many sectors; they spawn further innovation in application sectors, and they themselves are rapidly improving. As emphasized by Bresnahan and Trajtenberg (1995), the presence of a general-purpose technology gives rise to both vertical and horizontal externalities in the innovation process that can lead not just to underinvestment but also to distortions in the direction of investment, depending on the degree to which private and social returns diverge across different application sectors. Most notably, if there are "innovation complementarities" between the general purpose technology and each of the application sectors, lack of incentives in one sector can create an indirect externality that results in a system-wide reduction in innovative investment itself. While the private incentives for innovative investment in each application sector depend on its the market structure and appropriability conditions, that sector's innovation enhances innovation in the GPT itself, which then induces subsequent demand (and further innovation) in other downstream application sectors. These gains can rarely be appropriated within the originating sector. Lack of coordination between the GPT and application sectors, *as* well as across application sectors, is therefore likely to significantly reduce investment in innovation. Despite 6 these challenges, a reinforcing cycle of innovation between the GPT and a myriad of application sectors can generate a more systemic economy-wide transformation as the rate of innovation increases across all sectors. A rich empirical literature examining the productivity impacts of information technology point to *the* role of the microprocessor as a GPT as a way of understanding *the* impact of IT on the economy as a whole (among many others, Bresnahan and Greenstein (1995); Brynjolfsson and Hitt (1999); and Bresnahan, Brynjolfsson, and Hitt (2001)). Various aspects *of* artificial intelligence can certainly be understood as a GPT, and learning from examples such as the microprocessor are likely to be a useful foundation for thinking about both the magnitude of their impact on the economy, and associated policy challenges. A second conceptual framework for thinking about *AI* is the economics of research tools. Within the research sectors, some innovations open up new avenues of inquiry, or simply improve productivity "within the lab". Some of these advances appear to have great potential across a broad set of domains, beyond their initial application: as highlighted by Griliches (1957) in his classic studies of hybrid corn, some new research tools are inventions that do not just create or improve a specific product—instead they constitute a new way of creating new products, with much broader application. In Griliches' famous construction, the discovery of double-cross hybridization "was the invention of a method of inventing." (Hereinafter, "IMI".) Rather than being a means of creating a single a new corn variety, hybrid corn represented a widely applicable method for breeding many different new varieties. When applied to the challenge of creating new varieties optimized for many different localities (and even more broadly, to other crops) the invention of double-cross hybridization had a huge impact on agricultural productivity. One of the important insights to be gained from thinking about IMIs, therefore, is that the economic impact of some types of research tools is not limited to their ability to reduce the costs of specific innovation activities—perhaps even more consequentially they enable a new approach to innovation itself, by altering the "playbook" for innovation in the domains where the new tool is applied. For example, prior to the systematic *understanding* of the power of "hybrid vigor," a primary focus in agriculture had been improved techniques for self-fertilization (i.e., allowing for more and more specialized natural varietals over time). Once the rules governing hybridization (i.e., heterosis) were systematized, and the performance advantages of hybrid vigor 7 demonstrated, the techniques and conceptual approach for agricultural innovation was shifted, ushering in a long period of systematic innovation using these new tools and knowledge. Advances in machine learning and neural networks appear to have great potential as a research tool in problems of classification and prediction. These are both important limiting factors in a variety of research tasks, and, as exemplified by the Atomwise example, application of "learning" approaches to AI hold out the prospect of dramatically lower costs and improved performance in R&D projects where these are significant challenges. But as with hybrid corn, AI based learning may be more usefully understood as an IMI than as a narrowly limited solution to a specific problem. One the one hand, AI based learning may be able to substantially "automate discovery" across many domains where classification and prediction tasks play an important role. On the other, they may also "expand the playbook" is the sense of opening up the set of problems that can be feasibly addressed, and radically altering scientific and technical communities' conceptual approaches and framing of problems. The invention of optical lenses in the 17 th century had important direct economic impact in applications such as spectacles. But optical lenses *in* *the* form of microscopes and telescopes also had enormous and long-lasting indirect effects on the progress of science, technological change, growth, and welfare: by making very small or very distant objects visible for the first time, lenses opened up entirely new domains of inquiry and technological opportunity. Leung et al. (2016), for example, evocatively characterize *machine* learning as an opportunity to "learn to read the genome" *in* ways that human cognition and perception cannot. Of course, many research tools are neither IMIs nor GPTs, and their primary impact is to reduce the cost or enhance *the* quality of an existing innovation process. For example, in the pharmaceutical industry, new kinds of materials promise to enhance *the* efficiency of specific research processes. Other research tools can indeed be thought of as IMIs but are nonetheless relatively limited in application. For example, *the* development of genetically engineered research mice (such as the Oncomouse) is an IMI that *has* had a profound impact on the conduct and "playbook" of biomedical research, but has no obvious relevance to innovation in *areas* such as information technology, energy, or aerospace. The challenge presented by advances in AI is that they appear to be research tools that not only *have* *the* potential to change the method of innovation itself but also have implications across an extraordinarily wide range of fields. 8 Historically technologies with these characteristics—think of digital computing—have had large and unanticipated impacts across the economy and society in general. Mokyr (2002) points to the profound impact of IMIs that take the form not of tools per se, but innovations in the way research is organized and conducted, such as the invention of the university. GPTs that are themselves IMIs (or vice versa) are particularly complex phenomena, whose dynamics are as yet poorly understood or characterized. From a policy perspective, a further important feature of research tools is that it may be particularly difficult to appropriate their benefits. As emphasized by Scotchmer (1990), providing appropriate incentives for an upstream innovator that develops only the first "stage" of an innovation (such as a research tool) can be particularly problematic when contracting is imperfect and the ultimate application of the new products whose development is enabled by the upstream innovation is uncertain. Scotchmer and her co-authors emphasized a key point about a multi-stage research process: when the ultimate innovation that creates value requires multiple steps, providing appropriate innovation incentives are not only a question of whether *and* how to provide property rights in general, but also of how best to distribute property rights and incentives across the multiple stages of the innovation process. Lack of incentives for early- stage innovation can therefore mean that the tools required for subsequent innovation do not even get invented; strong early-stage property rights without adequate contracting opportunities *may* result in "hold-up" for later-stage innovators and so reduce the ultimate impact of the tool in terms of commercial application. The vertical research spillovers created by new research tools (or IMIs) are not just a challenge for designing appropriate intellectual property policy. 1 They are also exemplars *of* the core innovation externality highlighted by endogenous growth theory (Romer, 1990; Aghion and Howitt, 1992); a central source of underinvestment in innovation is *the* fact that the intertemporal spillovers from innovators today to innovators tomorrow cannot be easily captured. While tomorrow's innovators benefit from "standing on the shoulders of giants," their gains are not easily shared with their predecessors. This is not simply a theoretical idea: an increasing body of evidence suggests that research tools and the institutions that support their development and 1 Challenges presented by AI-enabled invention for legal doctrine and the patent process are beyond *the* scope of this essay. 9 diffusion play an important role in generating intertemporal spillovers (among others, Furman and Stern, 2011; Williams, 2014). A central insight of this work is that control—both *in* *the* form of physical exclusivity *as* well as *in* *the* form of formal intellectual property rights—over tools and data can shape both the level and direction of innovative activity, and that rules and institutions governing control over these areas has a powerful influence on the realized amount and nature of innovation. Of course, these frameworks cover only a subset *of* the key informational and competitive distortions that might arise when considering whether *and* how to provide optimal incentives for the type of technological change represented by some areas of AI. But these two areas in particular seem likely to be important for understanding the implications *of* the current dramatic advances in AI supported learning. We therefore turn in the next section to a brief outline of the ways *in* which AI is changing, with an eye towards bringing the framework here to bear on how we might outline a research agenda exploring the innovation policy challenges that they create. III. The Evolution of Artificial Intelligence: Robotics, Symbolic Systems, and Neural Networks In his omnibus historical account of AI research, Nilsson (2010) defines AI as "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment." His account details the contributions of multiple fields to achievements in AI, including but not limited to biology, linguistics, psychology and cognitive sciences, neuroscience, mathematics, philosophy and logic, engineering and computer science. And, of course, regardless of their particular approach, artificial intelligence research has been united by from the beginning by its engagement with Turing (1950), and his discussion of the possibility of mechanizing intelligence. Though often grouped together, the intellectual *history* of AI as a scientific and technical field is usefully informed by distinguishing between three interrelated but separate areas: robotics, neural networks, and symbolic systems. Perhaps the most successful line of research in the early years of AI—dating back to the 1960s—falls under the broad heading of symbolic 10 systems. Although early pioneers such as Turing had emphasized the importance of teaching a machine as one might a child (i.e., emphasizing AI as a learning process), the "symbol processing hypothesis" (Newell, Shaw, and Simon, 1958; Newell and Simon, 1976) was premised on the attempt to replicate the logical flow of human decision making through processing symbols. Early attempts to instantiate this approach yielded striking success in demonstration projects, such as the ability of a computer to navigate elements of a chess game (or other board games) or engage in relatively simple conversations with humans by following specific heuristics and rules embedded into a program. However, while research *based* on the concept of a "general problem solver" has continued to be an area of significant academic interest, and *there* have been periodic explosions of *interest* in *the* use of such approaches to assist human decision-making (e.g., in the context of early-stage expert systems to guide medical diagnosis), the symbolic systems approach has been heavily criticized for its inability to meaningfully impact real-world processes in a scalable way. It is of course possible that this field will see breakthroughs in the future, but it is fair to say that, while symbolic systems continues to be an area of academic research, it has not been central to the commercial application of AI. Nor is it at the heart of the recent reported advances in AI that are associated with the area of machine learning and prediction. A second influential trajectory *in* *AI* has been broadly in the area of robotics. While the concepts of "robots" as machines that can perform human tasks dates back at least to the 1940s, *the* field of robotics began to meaningfully flourish from the 1980s onwards through a combination of the advances in numerically controlled machine tools and *the* development of more adaptive but still rules-based robotics that rely on the active sensing of a known environment. Perhaps the most economically consequential *application* of AI to date has been in this area, with large scale deployment of "industrial robots" in manufacturing applications. These machines are precisely programmed to undertake a given task in a highly controlled environment. Often located in "cages" within highly specialized industrial processes (most notably automobile manufacturing), these purpose-built tools are perhaps more aptly described as highly sophisticated numerically controlled machines rather than as robots with significant AI content. Over the past twenty years, innovation in robotics has had an important impact on manufacturing and automation, most notably through the introduction of more responsive robots that rely on programmed response algorithms that can *respond* to a variety of stimuli. This 11 approach, famously pioneered by Rod Brooks (1990), focused the commercial and innovation orientation of AI away from the modeling of human-like intelligence towards providing feedback mechanisms that would allow for practical and effective robotics for specified applications. This insight led, among other applications, to the Roomba and to other adaptable industrial robots that could interact with humans such as Rethink Robotics' Baxter). Continued innovation in robotics technologies (particularly in the ability of robotic devices to sense and interact with their environment) may lead to wider application and adoption outside industrial automation. These advances are important, and the most advanced robots continue to capture public imagination when the term AI is invoked. But innovations in robotics are not, generally speaking, IMIs. The increasing automation of laboratory equipment certainly improves research productivity, but advances in robotics are not (yet) centrally connected to the underlying ways in which researchers themselves might develop approaches to undertake innovation itself across multiple domains. There are of course counterexamples to this proposition: robotic space probes have been a very important research tool in planetary science, and the ability of automated remote sensing devices to collect data at very large scale or in challenging environments may transform some fields of research. But robots continue to be used principally in specialized end-use "production" applications. Finally, a third stream of research that has been a central element of AI since its founding can be broadly characterized as a "learning" approach. Rather than being focused on symbolic logic, or precise sense-and-react systems, the learning approach attempts to create reliable and accurate methods for the prediction of particular events (either physical or logical) in the presence of particular inputs. The concept of a neural network has been particularly important in this area. A neural network is a program that uses a combination of weights and thresholds to translate *a* set of inputs into *a* set of outputs, measures the "closeness" of these outputs to reality, and then adjusts the weights it uses to narrow the distance between outputs and reality. In this way, neural networks can learn as they are fed more inputs (Rosenblatt, 1958; 1963). Over the course of the 1980s, Hinton and his co-authors further advanced the conceptual framework on which neural networks are based *through* *the* development of "back-propagating multi-layer" techniques that further enhance their potential for supervised learning. 12 After being initially heralded as having significant promise, *the* field of neural networks has come in and out of fashion, particularly within the United States. From the 1980s through the mid-2000s, their challenge seemed to be that there were significant limitations to the technology that could not be easily fixed by using larger training datasets or through the introduction of additional layers of "neurons." However, in the mid-2000s, a small number of new algorithmic approaches demonstrated *the* potential to enhance prediction through back propagation through multiple layers. These neural networks increased their predictive power as they were applied to larger and larger datasets, and were able to scale to an arbitrary level (among others, a key reference here is Hinton and Salakhutdinov (2006)). These advances exhibited a "surprising" level of performance improvement, notably in the context of the ImageNet visual recognition project competition pioneered by Fei-Fei Li at Stanford (Krizhevsky, Sutskever and Hinton, 2012). IV. How Might Different Fields within Artificial Intelligence Impact Innovation? Distinguishing between these three streams *of* AI is a critical first step towards developing *a* better *understanding* of how AI is likely *to* influence the innovation process going forward, since the three differ significantly in their potential to be either GPTs or IMIs—or both. First, though a significant amount of public discussion of *AI* *focuses* on the potential for AI to achieve super-human performance over a wide range of human cognitive capabilities, *it* *is* important to note that, at least so far, the significant advances *in* AI have not been *in* *the* form of the "general problem solver" approaches that were at the core of early work in symbolic systems (and that were the motivation for considerations of human reasoning such as the Turing test). Instead, recent advances in both robotics and in deep learning are by and large innovations that require a significant level of human planning and that apply to a relatively narrow domain of problem-solving (e.g., face recognition, playing Go, picking up a particular object, etc.) While it is of course possible that further breakthroughs will lead to a technology that can meaningfully mimic the nature of human subjective intelligence and emotion, the recent advances that have attracted scientific and commercial attention are well removed from these domains. 13 Second, though most economic and policy analysis of AI draws out consequences from the last two decades of automation to consider the future economic *impact* of AI (e.g., in job displacement for an ever-increasing number of tasks), *it* *is* important to emphasize that *there* is a sharp difference between the advances in robotics that were a *primary* focus of applications *of* AI research during the 2000s and the potential applications of deep learning which have come to the fore over the last few years. As we suggested above, current advances in robotics are by and large associated with applications that are highly specialized and that *are* focused on end-user applications rather than on the innovation process itself and these advances do not seem as of yet to have translated to a more generally applicable IMI. Robotics is therefore an area where we might *focus* *on* *the* impact of innovation (improved performance) and diffusion (more widespread application) in terms of job displacement versus job enhancement. We see limited evidence as yet of widespread applications of robotics outside industrial automation, or of the scale of improvements in *the* ability to sense, react to, and manipulate the physically environment that *the* use of robotics outside manufacturing probably requires. But there are exceptions: developments in the capabilities of “pick and place” robots and rapid progress in autonomous vehicles point to the possibility for robotics to escape manufacturing and become much more broadly used. Advances in robotics may well reveal this area of AI be a GPT, as defined by the classic criteria. Some research tools/IMIs based on algorithms have transformed the nature of research in some fields, but have lacked generality. These types of algorithmic research tools, based on a static set of program instructions, are a valuable IMI, but do not appear to have wide applicability outside a specific domain and do not qualify as GPTs. For example, while far from perfect, powerful algorithms to scan brain images (so-called functional MRI imaging) have transformed *our* *understanding* *of* the human brain, not only through the knowledge they have generated but also by establishing an entirely new paradigm and protocol for brain research. However, despite its role as a powerful IMI, fMRI lacks the type of general-purpose applicability that has been associated with the most important GPTs. *In* contrast, the latest advances in deep learning *have* *the* potential to be both a general-purpose IMI and a classic GPT. The following table summarizes these ideas: 14 General-Purpose Technology NO YES Invention of a Method of Invention NO Industrial Robots (e.g. Fanuc R2000) ‘Sense & React” Robots (e.g. Autonomous vehicles) YES Statically-coded Algorithmic Tools (e.g. fMRI) Deep Learning How might the promise of deep learning as a general-purpose IMI be realized? Deep learning promises to be an enormously powerful new tool that allows for the unstructured “prediction” of physical or logical events in contexts where algorithms based on a static set of program instructions (such as classic statistical methods) perform poorly. The development of this new approach to prediction enables a new approach to undertaking scientific and technical research. Rather than focusing on small well-characterized datasets or testing settings, it is now possible to proceed by identifying large pools of unstructured data which can be used to dynamically develop highly accurate predictions of technical and behavioral phenomena. In pioneering an unstructured approach to predictive drug candidate selection that brings together a vast array of previously disparate clinical and biophysical data, for example, Atomwise may fundamentally reshape the “ideas production function” in drug discovery. If advances in deep learning do represent the arrival of a general-purpose IMI, it is clear that there are likely to be very significant long-run economic, social, and technological consequence. First, as this new IMI diffuses across many application sectors, the resulting explosion in technological opportunities and increased productivity of R&D seem likely to generate economic growth that can eclipse any near-term *impact* *of* AI on jobs, organizations, and productivity. A more subtle implication of this point is that “past is not prologue”: even if automation over the recent past has resulted in job displacement (e.g., Acemoglu and Restrepo, 2017a), AI is likely to have at least as important an impact through *its* ability to enhance the potential for “new tasks” (as in Acemoglu and Restrepo, 2017b). Second, the arrival of a general-purpose IMI is a sufficiently uncommon occurrence that its impact could be profound for economic growth and its broader impact on society. There have 15 been only a handful of previous general-purpose IMIs and each of these has had an enormous impact not primarily through their direct effects (e.g., spectacles, in the case of the invention of optical lenses) but through their ability to reshape the ideas production function itself (e.g. telescopes and microscopes). It would therefore be helpful to understand the extent to which deep learning is, or will, causing researchers to significantly shift or reorient their approach in order to enhance research productivity (in the spirit of Jones (2009)). Finally, if deep learning does indeed prove to be a general-purpose IMI, it will be important to develop institutions and a policy environment that is conductive to enhancing innovation through this approach, and to do so in a way that promotes competition and social welfare. A central concern here may be the interplay between a key input required for deep learning—large unstructured databases that provide information about physical or logical events—and the nature of competition. While the underlying algorithms for deep learning are in the public domain (and can and are being improved on rapidly), the data pools that are essential to generate predictions may be public or private, and access to them will depend on organizational boundaries, policy and institutions. Because *the* performance of deep learning algorithms depends critically on the training data *that* they are created from, it may be possible, in a particular application area, for a specific company (either an incumbent or start-up) gain a significant, persistent innovation advantage through their control over data that is independent of traditional economies of scale or demand-side network effects. This “competition for the market” is likely to have several consequences. First, it creates incentives for duplicative racing to establish a data advantage in particular application sectors (say, search, autonomous driving, or cytology) *followed* by the establishment of durable barriers to entry that may be of significant concern for competition policy. Perhaps even more importantly, this kind of behavior could result in a balkanization of data within each sector, not only reducing innovative productivity within the sector, but also reducing spillovers back to the deep learning GPT sector, and to other application sectors. This suggests that the proactive development of institutions and policies that encourage competition, data sharing, and openness is likely to be an important determinant of economic gains from the development *and* application of deep learning. Our discussion so far has been largely speculative, and it would be useful to know whether our claim *that* deep learning may be both a general-purpose IMI and a GPT, while 16 symbolic logic and robotics are probably not, have any empirical basis. We turn in the next section to a preliminary examination of the evolution of AI as revealed by bibliometric data, with an eye towards answering this question. V.

Data This analysis draws upon two distinct datasets, one that captures *a* set of AI publications from Thompson Reuters Web of Science, and another that identifies *a* set of AI patents issued by the U.S. Patent and Trademark Office. In this section, we provide detail on the assembly of these datasets and summary statistics for variables in the sample. . As previously discussed, peer-reviewed and public-domain literature on AI points to the existence of three distinct fields within AI: robotics, learning systems and symbol systems, each comprised of numerous subfields. To track development of each of these using this data, we began by identifying the publications and patents falling into each of these three fields based on keywords. Appendix 1 lists the terms we used to define each field and identify the papers and patents belonging to it. . 2 *In* short, the robotics field includes approaches in which a system engages with and responds to environmental conditions; the symbolic systems field attempts to represent complex concepts through logical manipulation of symbolic representations, and the learning systems field processes data through analytical programs modeled on neurologic systems. Publication Sample and Summary Statistics Our analysis focuses on journal articles and book publications through the *Web* of Science from 1955 to 2015. We conducted a keyword search utilizing the keywords described in Appendix A (we tried several variants of these keywords and alternative algorithmic approaches but this did not result in a meaningful difference in the publication set). We are able to gather detailed information about each publication, including publication year, journal information, topical information, *as* well as author and institutional affiliations. 2 Ironically enough, we relied upon human intelligence rather than machine learning to develop this classification system and apply it to this data set. 17 This search yields 98,124 publications. We then code each publication into one of the three main fields of AI, as described above. Overall, relative to an initial dataset of 98,124, we are able to uniquely classify 95,840 publications as symbolic systems, learning systems, robotics, or “general” AI (we drop papers that involve combinations of these three fields). Table 1A reports the summary statistics for this sample. Of the 95,840 publication in the sample, 11,938 (12.5 percent) are classified as symbolic systems, 58,853 (61.4 percent) as learning and 20,655 (21.6 percent) as robotics, with the remainder being in the general field of “artificial intelligence.” To derive *a* better *understanding* of the factors that have shaped the evolution of AI, we create indicators for variables of interest including organization type (private versus academic), location type (US domestic versus International), and application type (computer science versus other application area, in addition to individual subject spaces, e.g. biology, materials science, medicine, physics, economics, etc.). We identify organization type as academic if the organization of one of the authors on the publication is an academic institution. 81,998 publications (85.5 percent) and 13,842 (14.4 percent) are produced by academic and private sector authors, respectively. We identify publication location as US domestic if one of the authors on the publication lists the United States as his or her primary location. 22,436 publications (25 percent of the sample) are produced domestically. We also differentiate between subject matter. 44 percent of the publications are classified as Computer Science, with 56 percent classified as other applications. Summary statistics on the other applications are provided in Table 2A. The other subjects with the largest *number* of publications in the sample include Telecommunications (5.5 percent), Mathematics (4.2), Neurology (3.8), Chemistry (3.7), Physics (3.4), Biology (3.4), and Medicine (3.1). Finally, we create indicator variables to document publication quality, including journal quality (top 10, top 25 and top 50 journals by impact factor 3) and a count variable for cumulative citation counts. Less than one percent of publications are in a top 10 journal with two percent and 10 percent in top 25 and top 50 journals. The average citation count for a publication in the sample is 4.9. 3 The rankings are collected from Guide2Research, found here: http://www.guide2research.com/journals/ 18 Patent Sample and Summary Statistics We undertake a similar approach for gathering a dataset of AI patents. We start with the public-use file of USPTO patents (Marco, Carley et al., 2015; Marco et al., 2015,), and filter the data in two ways. First, we assemble a subset of data by filtering the USPTO Historical Masterfile on the U.S. Patent Classification System (USPC) number. 4 Specifically, USPC numbers 706 and 901 represent “Artificial Intelligence” and “Robots,” respectively. Within USPC 706, there are numerous subclasses including “fuzzy logic hardware,” “plural processing systems,” “machine learning,” and “knowledge processing systems,” to name a few. We then use the USPC subclass to identify patents in AI fields of symbolic systems, learning systems and robotics. We drop patents prior to 1990, providing a sample of 7,347 patents through 2014. Second, we assemble another subset of AI patents by conducting a title search on patents, with the search terms being the same keywords used to identify academic publications in AI. 5 This provides an additional 8,640 AI patents. We then allocate each patent into an AI field by associating the relevant search term with one of the overarching fields. For example, a patent that is found through the search term “neural network,” is then classified as a “learning” patent. Some patents found through this search method will be duplicative of those identified by USPC search, i.e. the USPC class will be 706 or 901. We drop those duplicates. Together these two subsets create a sample of 13,615 unique AI patents. Summary statistics are provided in Table 1B. In contrast to the distribution of learning systems, symbolic systems and robotics in the publication data, the three fields are more evenly distributed in the patent data: 3,832 (28 percent) learning system patents, 3,930 (29 percent) symbolic system patents, and 5,524 (40 percent) robotics patents. The remaining patents are broadly classified only as AI. Using ancillary datasets to the USPTO Historical Masterfile, we are able to integrate variables of interest related to organization type, location, and application space. For example, 4 We utilized data from the Historical Patent Data Files. The complete (un-filtered) data sets from which we derived our data set are available here: https://www.uspto.gov/learning-and-resources/electronic-data-products/historical- patent-data-files 5 We utilized data from the Document ID Dataset that is complementary to Patent Assignment Data available on the USPTO website. The complete (un-filtered) data sets from which we derived our data set are available here: https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-assignment-dataset 19 Patent Assignment Data tracks ownership of patents across time. Our interest in this analysis relates to upstream innovative work, and for this reason, we capture the initial patent assignee by organization for each patent in our sample. This data enables the creation of indicator variables for organization type and location. We create an indicator for academic organization type by searching the name of the assignee for words relating to academic institutions, e.g. “University”, “College” or “Institution.” We do the same for private sector organizations, searching for “corp”, “business”, “inc”, or “co”, to name a few. We also search for the same words or abbreviations utilized in other languages, e.g. “S.p.A.” Only seven percent of the sample is awarded to academic organizations, while 91 percent is awarded to private entities. The remaining patents are assigned to government entities, e.g. U.S. Department of Defense. Similarly, we create indicator variables for patents assigned to U.S. firms and international firms, *based* on the country of the assignee. The international firm data can also be more narrowly identified by specific country (e.g. Canada) or region (e.g. European Union). 59 percent of our patent sample is assigned to U.S. domestic firms, while 41 percent is assigned to international firms. Next to the United States, firms from non-Chinese, Asian nations account for 28 percent of patents in the sample. Firms from Canada are assigned 1.2 percent of the patents, and firms from China, 0.4 percent. Additionally, the USPTO data includes NBER classification and sub-classification for each patent (Hall, Jaffe and Trajtenberg (2001); Marco, Carley, et al., (2015)). These sub-classifications provide some granular detail about the application sector for which the patent is intended. We create indicator variables for NBER sub-classifications related to chemicals (NBER sub-class 11, 12, 13, 14, 15, 19), communications (21), computer hardware and software (22), computer science peripherals (23), data and storage (24), business software (25), medical fields (31, 32, 33, and 39), electronics fields (41, 42, 43, 44, 45, 46, and 49), automotive fields (53, 54, 55), mechanical fields (51, 52, 59), and other fields (remaining). The vast majority of these patents (71 percent) are in NBER subclass 22, Computer Hardware and Software. Summary Statistics of the distribution of patents across application sectors are provided in Table 2B. 20 VI. Deep Learning as a GPT: An Exploratory Empirical Analysis These data allow us to begin examining the claim that the technologies of deep learning may be the nucleus of a general-purpose invention for the method of invention. We begin in Figures 1A and 1B with a simple description of the evolution over time of the three main fields *identified* in the corpus of patents and papers. The first insight is that the overall *field* *of* AI has experienced sharp growth since 1990. While there are only a small handful of papers (less than a hundred per year) at the beginning of the period, each of the three fields now generates more than a thousand papers per year. *At* *the* same time, *there* is a striking divergence in activity across fields: each start from a similar base, but *there* is a steady increase in the deep learning publications relative to robotics and symbolic systems, particularly after 2009. Interestingly, at least through the end of 2014, there is more similarity in the patterns for all three fields in terms of patenting, with robotics patenting continuing to hold a lead over learning and symbolic systems. However, there does seem to be an acceleration of learning- oriented patents in the last few years of the sample, and so there may be a relative shift towards learning over the last few years which will manifest itself over time as publication and examination lags work their way through. Within the publication data, there are striking variations across geographies. Figure 2A shows the overall growth in learning publications for the US versus rest-of-world, and Figure 2B maps the fraction of publications within each geography that are learning related. In the US on learning is far more variable. Prior to 2000 the US has a roughly equivalent share of learning related publications, but the US then falls significantly behind, only catching up again around 2013. This is consistent with the suggestion in qualitative histories of AI that that learning research *has* had a “faddish” quality in the US, with the additional insight that the rest of the world (notably Canada) seems to have taken advantage of this inconsistent focus in the United States to develop capabilities and comparative advantage in this field. With these broad patterns in mind, we turn to our key empirical exercise: whether in the late 2000s deep learning shifted more towards “application-oriented” research than either robotics or symbolic systems. We begin in Figure 3 with a simple graph that examines the *number* of publications over time (across all three fields) in computer science journals versus application-oriented outlets. While there has actually been a stagnation (even a small decline) in 21 the overall number of AI publications in computer science journals, there has been a dramatic increase in the number of AI-related publications in application-oriented outlets. By the end of 2015, we estimate that nearly 2/3 of all publications in AI were in fields beyond computer science. In Figure 4 we then look at this division by field. Several patterns are worthy of note. First, as earlier, we can see the relative growth through 2009 of publications in learning versus the two other fields. Also, consistent with more qualitative accounts of the fields, we see the relative stagnation of symbolic systems research relative to robotics and learning. But, after 2009, *there* is a significant increase in application publications in both robotics and learning, but that the learning boost is both steeper and more long-lived. Over the course of just seven years, learning-oriented application publications more than double in number, and now represent just under 50% of all AI publications. 6 These patterns are if anything even more striking if one disaggregates them by the geographic origin of the publication. In Figure 5, we at rates of publication in computer science versus applications for the US versus rest-of-world. The striking upward swing in AI application papers that begins in 2009 turns out to be overwhelmingly driven by publications ex US, though US researchers begin a period of catch-up at an accelerating pace towards the final few years of the sample. Finally, we look at how publications have varied across application sectors over time. In Table 3, we examine the *number* of publications by application field in each of the *three* areas of AI across two three-year cohorts (2004-2006 and 2013-2015). There are a number of patterns of interest. First, and most importantly, in a range of application fields including medicine, radiology and economics, *there* is a large relative increase in learning-oriented publications relative to robotics and symbolic systems. A number of other sectors, including neuroscience and biology, realize a large increase in both learning-oriented research *as* well as other AI fields. There are also some more basic fields such as mathematics that have experienced a relative decline in publications (indeed, learning-oriented publications in mathematics experienced a 6 The precise *number* of publications for 2015 are estimated from the experience of the first nine months (the *Web* of Science data run through September 30, 2015). We apply a linear multiplier for the remaining three months (i.e., estimating each category by 4/3). 22 small absolute decline, a striking different relative to most other fields in the sample). Overall, though it would be useful to identify more precisely the type of research that is being conducted and what is happening at *the* level of particular subfields, these results are consistent with our broader hypothesis that, alongside the overall growth of AI, learning-oriented research may represent a general-purpose technology that is now beginning to be exploited far more systematically across a wide range of application sectors. Together, these preliminary findings provide some direct empirical evidence for at least one of our hypotheses: learning-oriented AI seems to have some of the signature hallmarks of a general-purpose technology. Bibliometric indicators of innovation show *that* it is rapidly developing, and is being applied in many sectors—and these application sectors themselves include some of the most technologically dynamic parts of the economy. This preliminary analysis does not trace out the important knowledge spillovers between innovation in the GPT and innovation and application sectors, but it is probably far too early to look for evidence of this. VII. Deep Learning as a General-Purpose Invention in the Method of Invention: Considerations for Organizations, Institutions and Policy With these results in mind, we now consider the potential implications for innovation and innovation policy if deep learning is indeed a general-purpose technology (GPT) and/or a general-purpose invention in the method of invention (IMI). If deep learning is merely a GPT, it is likely to generate innovation across a range of applications (with potential for spillovers both back to the learning GPT *and* also to other application sectors) but will not itself change the *nature* of the innovation production function. If it is also a general purpose IMI, we would expect it to have an even larger impact on economy-wide innovation, growth, and productivity as dynamics play out—and to trigger even more severe short run disruptions of labor markets and the internal structure of organizations. Widespread use of deep learning as a research tool implies a shift towards investigative approaches that use large datasets to generate predictions for physical and logical events that have previously resisted systematic empirical scrutiny. These data are likely to have three 23 sources: prior knowledge (as in the case of “learning” of prior literatures by IBM’s Watson), online transactions (e.g., search or online purchasing behavior) and physical events (e.g., the output from various types of sensors or geolocation data) What would this imply for the appropriate organization of innovation, the institutions we have for training and conducting research over time, and for policy, particularly as we think about private incentives to maintain proprietary datasets and application-specific algorithms? The Management and Organization of Innovation Perhaps most immediately, the rise of general-purpose predictive analytics using large datasets seems likely to result in a substitution towards capital and away from labor *in* the research production process. Many types of R&D and innovation more generally are effectively problems of labor-intensive search with high marginal cost per search (Evenson and Kislev, 1975, among others). The development of deep learning holds out the promise of sharply reduced marginal search costs, inducing R&D organizations to substitute away from highly- skilled labor towards fixed cost investments in AI. These investments are likely to improve performance in existing “search intensive” research projects, *as* well as to open up new opportunities to investigate social and physical phenomena that have previously been considered intractable or even as beyond *the* domain of systematic scientific and empirical research. It is possible that *the* ability to substitute away from specialized labor and towards capital (that in principle could be rented or shared) may lower the “barriers to entry” in certain scientific or research fields—particularly those in which the necessary data and algorithms are freely available—while erecting new barriers to entry in other areas (e.g. by restricting access to data and algorithms). As of yet, there are few if any organized markets for “trained” research tools or services based on deep learning, and few standards to evaluate alternatives. Our analysis suggests that *the* development of markets for shared AI services and the widespread availability of relevant data may be a necessary precursor to the broad adoption and dissemination of deep learning. *At* *the* *same* time, the arrival of this new research paradigm is likely to require a significant shift in the management of innovation itself. For example, it is possible that the democratization of innovation will also be accompanied by a lack of investment by individual researchers in specialized research skills and specialized expertise in any given area, reducing the 24 level of theoretical or technical depth in the work force. This shift away from career-oriented research trajectories towards *the* ability to derive new findings based on deep learning may undermine long-term incentives for breakthrough research that can only be conducted by people who are at the research frontier. There is also the possibility that the large scale replacement of skilled technical labor *in* the research sector by AI will "break science" in some fields by disrupting the career ladders and labor markets that support the relatively long periods of training and education required in many scientific and technical occupations. Finally, it is possible *that* deep learning will change the nature of scientific and technical advance itself. Many fields of science and engineering are driven by a mode of inquiry that focuses on identifying a relatively small number of causal drivers of underlying phenomena built upon an underlying theory (the parsimony principle as restated by Einstein states that theory should be "as simple as possible but no simpler.") However, deep learning offers an alternative paradigm *based* on *the* *ability* to predict complex multi-causal phenomena using a "black box" approach that abstracts away from underlying causes but that does allow for a singular prediction index that can yield sharp insight. De-emphasizing the understanding of causal mechanisms and abstract relationships may come at a cost: many major steps forward in science involve *the* ability to leverage *an* understanding of "big picture" theoretical structure to make sense of, of recognize the implications of, smaller discoveries. For example, it is easy to imagine *a* deep learning system trained on a large amount of x-ray diffraction data quickly "discovering" the double helix structure of DNA at very low marginal cost, but it would likely require human judgment and insight about a much broader biological context to notice that the proposed structure suggests a direct mechanism for heredity. Innovation and Competition Policy and Institutions A second area of impact, beyond the organization of individual research projects or the nature of what counts as "science" in a particular field, will be on the appropriate design and governance of institutions governing the innovation process. Three implications stand out. First, as discussed above, research over the past two decades has emphasized the important role played by institutions that encourage cumulative knowledge production through low-cost independent access to research tools, materials and data (Furman and Stern, 2012; Murray, et al, 2015). However to date there has only been a modest level of attention to the 25 questions of transparency and replicability within the deep learning community. Grassroots initiatives to encourage openness organized through online hubs and communities are to be welcomed. But it is useful to emphasize that there is likely to be a significant gap between the private and social incentives to share and aggregate data—even among academic researchers or private sector research communities. One implication of this divergence may be that to the degree any single research result depends on the aggregation of data from many sources, it will be important to develop rules of credit and attribution, *as* well as to develop mechanisms to replicate the results. This implies that it will be particularly important to pay attention to the design and enforcement of formal intellectual property rights. On the one hand it will be important to think carefully about the laws that currently surround the ownership of data. Should the data about e.g. my shopping and travel behavior belong to me or to the search engine or ride sharing company that I use? Might consumers have a strong collective interest in ensuring that these data (suitably blinded, of course) are in the public domain, so that many companies can use them in the pursuit of innovation? On the other, *the* advent of deep learning has significant implications for the patent system. Though there has so far been relatively little patenting of deep learning innovations, historical episodes such as the discovery and attempted wholesale patenting of express sequence tags and other kinds of genetic data suggests that breakthroughs in research tools—often combined with a lack of capacity at patent offices and conflicting court decisions—can result in long periods of uncertainty that has hampered the issuing of new patents, and this in turn has led to lower research productivity and less competition. Deep learning also presents difficult questions of legal doctrine for patent systems that have been built around *the* idea of creative authors and inventors. For example, "inventorship" has a specific meaning in patent law, with very important implications for ownership and control of the claimed invention. Can an AI system be an inventor in the sense envisaged by the drafters of the US Constitution? Similarly, standards for determining the size of the inventive step required to obtain a patent are driven by a determination of whether the claimed invention would or would not be obvious to a "person having ordinary skill in the art." Who this "person" might be, and what constitutes "ordinary 26 skill" in an age of deep learning systems trained on proprietary data, are questions well beyond *the* scope of this essay. *In* addition to these traditional innovation policy questions, the prospect for deep learning raises a wide variety of other issues, including issues relating to privacy, the potential for bias (deep learning has been found to reinforce stereotypes already present in society), and consumer protection (related to *areas* such as search, advertising, and consumer targeting and monitoring). The key is that, to the extent *that* deep learning is general-purpose, the issues that arise across each of these domains (and more) will play out across a wide variety of sectors and contexts and at a global rather than local level. Little analysis has been conducted that can help design institutions that will be responsive at *the* level of application sectors that also internalize the potential issues that may arise with *the* *fact* *that* deep learning is likely to be a GPT. Finally, the broad applicability of deep learning (and possibly robotics) across many sectors is likely to engender a race within each sector to establish a proprietary advantage that leverages these new approaches. As such, the arrival of deep learning raises issues for competition policy. In each application sector, there is the possibility that firms that are able to establish an advantage at an early stage, and in doing so position themselves to be able to generate more data (about their technology, about customer behavior, about their organizational processes) will be able to erect a deep-learning-driven barrier to entry that will ensure market dominance over at least the medium term. This suggests that rules ensuring data accessibility are not only a matter of research productivity or aggregation, but also speak to *the* potential to guard against lock-in and anticompetitive conduct. At the present moment there seem to be a large number of individual companies attempting to take advantage of AI across a wide variety of domains (e.g., there are probably more than 20 firms engaging in significant levels of research in autonomous vehicles, and no firm has yet to show a decisive advantage), but this high level of activity likely reflects an expectation for the prospects for significant market power in the future. Ensuring *that* deep learning does not enhance monopolization and increase barriers to entry across a range of sectors will be a key topic going forward. 27 VIII. Concluding Thoughts The purpose of this exploratory essay has not been *to* provide a systematic account or prediction of the likely *impact* *of* AI on innovation, nor clear guidance for policy or the management of innovation. Instead, our goal has been to raise a specific possibility—that deep learning represents a new general-purpose invention of a method of invention—and to draw out some preliminary implications of that hypothesis for management, institutions, and policy. Our preliminary analysis highlights a few key ideas that have not been central to the economics and policy discussion so far. First, at least from the perspective of innovation, it is useful to distinguish between the significant and important advances in fields such as robotics from the potential of a general-purpose method of invention based on application of multi- layered neural networks to large amounts of digital data to be an "invention in the method of invention". Both the existing qualitative evidence and our preliminary empirical analysis documents a striking shift since 2009 towards *deep* learning based application-oriented research that is consistent with this possibility. Second, and relatedly, the prospect of a change in the innovation process raises key issues for a range of policy and management areas, ranging from how to evaluate this new type of science to the potential for prediction methods to induce new barriers to entry across a wide range of industries. Proactive analysis of the appropriate private and public policy responses towards these breakthroughs seems like an extremely promising area for future research. 28 REFERENCES Aghion, P. and P. Howitt (1992) "A Model of Growth Through Creative Destruction," Econometrica, 60(2), 323-251. Bresnahan, T., E. Brynjolfsson, and L. Hitt (2002) "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," The Quarterly Journal of Economics, 117(1), 339-376. Bresnahan, T. and S. Greenstein (1999) "Technological Competition and the Structure of the Computer Industry," Journal of Industrial Economics, 47(1), 1-40. Bresnahan, T. and M. Trajtenberg (1995) "General Purpose Technologies 'Engines of Growth'?" Journal of Econometrics, 65 (1995) 83-108. Brooks, R. (1990) "Elephants Don't Play Chess," Robotics and Autonomous Systems, 6, 3-15. Brooks, R. (1991) "Intelligence Without Representation," Artificial Intelligence, 47, 139-159. Brynjolfsson, E. and K. McElheran (2017) "The Rapid Adoption of Data-Driven Decision-Making," American Economic Review, 106(5), 133-139 Griliches, Z. (1957) "Hybrid Corn: An Exploration in the Economics of Technological Change," Econometrica, 25(4), 501-522. Henderson, R. and K. Clark (1990) "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," Administrative Science Quarterly, 35(1), 9-30. Krizhevsky, A., I. Sutskever, G. Hinton (2012) "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing, 25, MIT Press. Leung, M.K.K., A. Delong, B. Alipanahi, and B.J. Frey (2016) "Machine Learning in Genomic Medicine: A Review of Computational Problems and Data Sets," *Proceedings* of the IEEE, 104(1): 176-197. Marco, A., A. Myers, S. Graham, P. D'Agostino, and K. Apple (2015) "The USPTO Patent Assignment Dataset: Descriptions and Analysis," USPTO Working Paper No. 2015-02, 1-53. Marco, A., M. Carley, S. Jackson and A. Myers (2015) "The USPTO Historical Patent Data Files," USPTO Working Paper No. 2015-01, 1-57. Minsky, M. (1961) "Steps Toward Artificial Intelligence," *Proceedings* of the IRE, 8-30. Mokyr, J (2002) Gifts of Athena, Princeton University Press. Nilsson, N. (2010), The Quest for *Artificial* Intelligence: A History of Ideas and Achievements, Cambridge University Press. Romer, P. (1990) "Endogenous Technological Change," Journal of Political Economy, 98(5), S71-S102. Rosenberg, N. and M. Trajtenberg (2004) "A General Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States," Journal of Economic History, 61(1), 61-99. 29 Rumelhart, D., G. Hinton, and R. Williams (1986) "Learning Internal Representations by Error Propagation," in J. McClelland and D. Rumelhart (editors), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models, MIT Press, 7-57. Scotchmer, S. (1991) "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," Journal of Economic Perspectives, 5(1), 29-41. Turing, A. (1950) "Computing Machinery and Intelligence," Mind, 59, 433-460. Wallach, I. Dzamba, M. and Heifels, A. "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery." arXiv:1510.02855 [cs.LG] Williams, H. .(2013) "Intellectual Property Rights and Innovation: Evidence

from the Human Genome,”, Journal of Political Economy, 121(1): 1-27 30 Table 1A: Publication Data Summary Statistics Mean Std. Dev. Min Max Publication Year 2007 6.15 1990 2015 Symbolic Systems .12 .33 0 1 Learning Systems .61 .48 0 1 Robotics .21 .41 0 1 Artificial Intelligence .06 .23 0 1 Computer Science .44 .50 0 1 Other Applications .56 .50 0 1 US Domestic .25 .43 0 1 International .75 .43 0 1 Observations 95840 Table 1B: Patent Data Summary Statistics Mean Std. Dev. Min Max Application Year 2003 6.68 1982 2014 Patent Year 2007 6.98 1990 2014 Symbolic Systems .29 .45 0 1 Learning Systems .28 .45 0 1 Robotics .41 .49 0 1 Artificial Intelligence .04 .19 0 1 Computer Science .77 .42 0 1 Other Applications .23 .42 0 1 US Domestic Firms .59 .49 0 1 International Firms .41 .49 0 1 Org Type Academic .07 .26 0 1 Org Type Private .91 .29 0 1 Observations 13615 31 Table 2A: Distribution of Publications across Subjects Mean Std. Dev. Biology .034 .18 Economics .028 .16 Physics .034 .18 Medicine .032 .18 Chemistry .038 .19 Mathematics .042 .20 Materials Science .029 .17 Neurology .038 .19 Energy .015 .12 Radiology .015 .12 Telecommunications .055 .23 Computer Science .44 .50 Observations 95840 Table 2B: Distribution of Patents across Application Sectors Mean Std. Dev. Chemicals .007 .08 Communications .044 .20 Computer Hardware and Software .710 .45 Computer Peripherals .004 .06 Data and Storage .008 .09 Business software .007 .09 All Computer Science .773 .42 Medical .020 .14 Electronics .073 .26 Automotive .023 .15 Mechanical .075 .26 Other .029 .16 Observations 13615 32 Table 3: Publications Across Sectors, by AI Field, 2004-2006 versus 2013-2015 BiologyEconomicsPhysicsMedicineChemistryMathMaterials Neuro. Energy Radiology Telecom. CompSci 2004-20062582923432313254172092711729429138889 2013-20156004233885164904144299702721864044582 % growth133%45%13%123%51%-1%105%258%58%98%39%18% 2004-200633105269244536316476531431 2013-2015651212283928022513918254011322 % growth97%20%135%20%283%78%525%348%200%-47%-39%-8% 2004-20069386896139543235158251827 2013-201510510125841496010173225688125 % growth13%25%84%-13%7%11%216%109%47%-32%73%36% Learning Systems Robotics Symbol Systems 33 Table 4: Herfindahl-Hirschman Index for Application Sectors Application AÓÒ ! G}H-I]G-H•H-I=Hb } 6†VÖ–6 Â Applications 153. Communications 140.87 Hardware and Software 86.99 Computer Science Peripherals 296 Data and Storage 366.71 Computer Science Business Models 222 Medical Applications 290.51 Electronic Applications 114.64 Automotive Applications 197.03 Mechanical Applications 77.51 Other 129.20 34 Figure 1A: Publications by AI field over Time Figure 1B: Patents by AI field over Time 0 1000 2000 3000 4000 5000 1990 1995 2000 2005 2010 2015 pubyear Learning Systems Symbolic Systems Robotics 0 200 400 600 800 199019952000200520102015 pat_year Learning SystemsSymbol Systems Robotics 35 Figure 2A: Academic Institution Publication Fraction by AI Field Figure 2B: Fraction of Learning Publications by US versus World 0 1000 2000 3000 4000 199019952000200520102015 pubyear U.S.A.International .3 .4 .5 .6 .7 (mean) learning 199019952000200520102015 pubyear Share of U.S. AI pubs in Learning Share of International AI pubs in Learning 36 Figure 3: Publications in Computer Science versus Application Journals Figure 4: Publications in Computer Science versus Application Journals, by AI Field 0 1000 2000 3000 4000 5000 All 199019952000200520102015 pubyear CompSci PublicationsApplication Publications 0 1000 2000 3000 199019952000200520102015 pubyear Learning (CS)Robotics (CS)Symbol (CS) Learning (Apps.)Robotics (Apps.)Symbol (Apps.) 37 Figure 5: Learning Publications in Computer Science versus Applications, By US versus ROW 0 500 1000 1500 2000 199019952000200520102015 pubyear U.S.A. (CS)International (CS) U.S.A. (Apps.)International (Apps.) 38 Appendix A Appendix Table 1: Artificial Intelligence Keyword Allocation