Artificial Interdisciplinarity:
Artificial Intelligence for Research on Complex Societal Problems

Seth D. Baum
Global Catastrophic Risk Institute
https://sethbaum.com * https://gcri.org

Philosophy & Technology, vol. 34, no. S1 (November 2021), p.45-63
DOI 10.1007/s13347-020-00416-5. This version 22 February 2022

Abstract

This paper considers the question: In what ways can artificial intelligence assist with interdisciplinary research for addressing complex societal problems and advancing the social good? Problems such as environmental protection, public health, and emerging technology governance do not fit neatly within traditional academic disciplines and therefore require an interdisciplinary approach. However, interdisciplinary research poses large cognitive challenges for human researchers that go beyond the substantial challenges of narrow disciplinary research. The challenges include epistemic divides between disciplines, the massive bodies of relevant literature, the peer review of work that integrates an eclectic mix of topics, and the transfer of interdisciplinary research insights from one problem to another. Artificial interdisciplinarity already helps with these challenges via search engines, recommendation engines, and automated content analysis. Future "strong artificial interdisciplinarity" based on human-level artificial general intelligence could excel at interdisciplinary research, but it may take a long time to develop and could pose major safety and ethical issues. Therefore, there is an important role for intermediate-term artificial interdisciplinarity systems that could make major contributions to addressing societal problems without the concerns associated with artificial general intelligence.

Keywords: artificial intelligence; interdisciplinary research; automated content analysis; artificial general intelligence

1. Introduction

This paper explores the question of how artificial intelligence (AI) can be of value to interdisciplinary research (IDR) aimed at addressing major societal problems such as public health, environmental protection, and emerging technology governance. These problems are highly complex and multifaceted and do not fit neatly into traditional academic disciplines. For example, the study of global warming integrates environmental science, social science, policy, energy engineering, and more. This extreme breadth makes IDR a very difficult cognitive challenge for even the most capable human researchers. If AI can help meet this challenge, that could be of considerable value for addressing many major societal problems.

This paper is especially interested in the possibilities for near- to intermediate-term AI systems that use existing AI techniques or relatively straightforward extensions of them. These would ideally be systems that computer scientists could work toward right now, and that would also be well short of strong AI, i.e., human-level or super-human-level artificial general intelligence (AGI). A central question of this paper is: *Can we imagine tractable, real-world AI system designs that would make significant contributions to understanding and addressing complex and important societal problems, but without posing the major risks and ethical issues* *associated with strong AI?* I humbly acknowledge that I do not know the answer to this question, but I do know that it is a worthy question to ask, and I can lay out some relevant considerations. That is the aim of this paper.

To streamline the discussion, let us introduce the following term:

> *Artificial interdisciplinarity* (A-ID): Artificial intelligence that performs interdisciplinary research or supports other agents in the performance of interdisciplinary research.

Some emphasis is warranted on the clause "or supports other agents in the performance of". Per this definition, an AI system that human researchers use to perform IDR counts as A-ID. Given this wider definition, it follows that there is A-ID in active use right now. Likewise, ideas for new forms of A-ID can leverage human collaboration. To make a valuable contribution to addressing societal problems, A-ID only needs to help with some aspect of IDR; it does not need to be able to complete IDR projects with no human collaboration. Indeed, it is plausible that an A-ID system that could succeed at the full range of IDR tasks would need to be strong AI.

As a scholarly contribution, this paper sits at the intersection of literatures on IDR and both near-term and long-term AI, as well as the nascent concept of intermediate-term AI. Regarding IDR, there is an active line of research on the cognitive challenges faced by human interdisciplinary researchers (Bracken and Oughton 2006; Keestra 2017; MacLeod 2018). To the best of my knowledge, this literature has not yet considered the current or potential future role of AI. Regarding near-term AI, an extensive literature applies existing AI tools to support the research process. Most of this is oriented toward narrow disciplinary research, but some studies involve IDR (Nunez-Mir et al. 2017; Tuhkala et al. 2018). There is also a more general, extensive, and longstanding use of computers in IDR, especially in the physical sciences (Crease 2017), though that is beyond the scope of this paper. Regarding long-term AI, some studies consider the prospect of superintelligent AI designed as "oracles" to answer a wide range of questions that humans might have, some of which are in the domain of IDR (Armstrong et al. 2012; Yampolskiy 2012). Each of these lines of research is discussed in greater detail later in the paper. Finally, recent research has argued that intermediate-term AI has gone overlooked relative to near-term and long-term AI (Parson et al. 2019a; 2019b; Baum 2020). This paper provides dedicated attention to intermediate-term AI and concurs that this time period merits attention.

The paper also seeks to encourage AI research to orient itself toward addressing important societal problems and to provide some direction for this. At present, much of AI research is instead oriented toward more intellectual goals of expanding the capacity of AI systems or toward profitable activities of technology corporations. As is well-known in a lot of IDR, intellectual progress and profit do not necessarily bring progress on societal problems, and in some cases can make the problems worse, for example intellectual progress on dangerous technologies and profitable activity that pollutes the environment. IDR is also not necessarily good for society and can be oriented toward a wide range of ends including intellectual progress and profit. Furthermore, even when IDR is focused on addressing societal problems, it is still not a panacea for them—the existence of literature addressing societal problems does not on its own ensure that the problems actually get addressed. That said, IDR is vital for addressing complex societal problems, and this is the primary focus of IDR today. The paper therefore encourages A-ID research to join forces with the extensive communities of IDR focused on addressing societal problems.

I approach the topic of AI and IDR as a veteran interdisciplinary researcher with some knowledge of both AI and the meta-study of IDR. I am not a computer scientist, and so I can speak with less confidence on what the AI options for IDR may be. Therefore, this paper seeks to lay out some general considerations and to hopefully serve as a prompt for an interdisciplinary conversation among computer scientists, scholars of IDR, and others with relevant perspectives. As this paper explains, this sort of interdisciplinary conversation can be difficult, but it is vital for making progress on a wide range of cross-cutting topics.

In keeping with much of the literature on IDR, this paper is written in terms of "societal problems" instead of the concept of "social good" that frames this special issue. These two terms are broadly compatible, in that addressing societal problems is one major way to advance the social good. IDR is commonly used to make progress on specific problems or issues, and is perhaps less commonly used to understand or advance the good in a more abstract or general sense. The challenge of understanding the complex and distinctive features of many specific societal problems is a primary focus of IDR.

As the first dedicated discussion of A-ID, the paper provides a broad discussion. Section 2 introduces IDR and outlines its cognitive challenges. Sections 3-5 discuss A-ID over the near-term, intermediate-term, and long-term. Section 6 discusses issues raised by A-ID. Section 7 concludes with general thoughts on the role of AI in addressing societal problems.

Sections 3-5 are organized around time scales because each time scale poses distinct issues for A-ID. Near-term AI can be defined as AI that currently exists or could exist in the near future via relatively straightforward extensions of current AI systems. Attention to near-term A-ID includes description of existing A-ID systems and prospects for extending them. Intermediate-term AI goes beyond what is already feasible or now in development, but it stops short of the more extreme forms of AI that could exist over the long-term. This paper defines long-term AI as AI with capabilities that equal or exceed human cognition; terms for this include strong AI, human-level AI, artificial general intelligence, ultraintelligence, and superintelligence. Discussions of AI are often divided between the near term and the long term, with some debate over which is more important (Baum 2018; Cave and Ó hÉigeartaigh 2019; Prunkl and Whittlestone 2020). Less attention has been paid to AI over intermediate time scales (Parson et al. 2019a; 2019b), though this paper finds that A-ID at all time scales can be important.

2. Interdisciplinary Research and its Cognitive Challenges

IDR has been defined in a variety of ways (Klein 2017). One notable definition is provided in the U.S. National Academies report *Facilitating Interdisciplinary Research*:

> Interdisciplinary research (IDR) is a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice (Committee on Facilitating Interdisciplinary Research and Committee on Science, Engineering, and Public Policy 2005, p.2).

Interdisciplinarity is often compared with related concepts, especially multidisciplinarity and transdisciplinarity. To generalize, multidisciplinary research is commonly understood to be the closest to disciplinary research, consisting of research that has input from multiple disciplines without integrating them; IDR takes an additional step away from disciplinary research by also integrating insights from across disciplines; transdisciplinary research is the furthest from disciplinary research by transcending disciplinary boundaries altogether, as if they did not exist, and often also transcending boundaries between academic and non-academic sources of knowledge. These distinctions are fuzzy and contested (Klein 2017) but suffice for purposes of this paper. Note that some of the research described in this paper may be more precisely classified as transdisciplinary, though the distinction between interdisciplinarity and transdisciplinarity is not crucial for this paper and the paper will use the two terms more or less interchangeably.

IDR has traditionally had two somewhat distinct motivations. See Bernstein (2015) for an intellectual history. One motivation is intellectual, seeking the understanding of cross-cutting topics and the synthesis of knowledge accumulated in (and beyond) the disciplines. An especially ambitious form of this motivation, associated mainly with transdisciplinarity, is for unifying all of the various strands of human knowledge and intellectual activity. The concept of consilience as developed by Wilson (1998) is one work in this direction. While intellectually motivated IDR is commonplace, IDR is perhaps more commonly motivated by practical societal problems. Environmental problems have been a longstanding focus, given their societal importance and multifaceted nature, involving both social and ecological systems. More generally, IDR often—but not always–addresses social and policy issues associated with science and technology; this includes social and policy issues raised by AI.

A different type of motivation for IDR is the complexity of many important subjects of study. Complexity can be defined in a variety of ways; one definition proposed in the context of IDR is that complexity is an attribute of systems whose components—which may themselves be systems–interact in predominantly nonlinear ways (Newell 2001). I would add that while this definition's emphasis on nonlinear systems has a mathematical tone, the interconnections found in IDR are often best understood in more qualitative terms. Regardless, the definition points to the distinction between multidisciplinarity and interdisciplinarity. Many research subjects do not fit neatly within any one academic discipline, but research on these subjects does not necessarily require integration across disciplines. For example, oceanography and battery technology are both relevant to the study of global warming, but oceanographers and battery engineers typically do not need to consider each others' work when doing their own. (This example is from a critique of IDR by Jacobs 2013, p.130-131.) This disconnected work would classify as multidisciplinary but not interdisciplinarity. As a contrasting example, global warming policy may affect industrial activity, which affects greenhouse gas emissions, which affect climate patterns, which affect natural hazards, which affect human welfare. Interdisciplinary global warming policy research may consider these complex interconnections to assess the overall merits of a policy idea.

All research, whether interdisciplinary or disciplinary, has social as well as intellectual dimensions. Indeed, one argument in favor of disciplines is that they provide a useful structure for organizing populations of human researchers within universities (Jacobs 2013). Interdisciplinary research centers serve a similar social purpose. Likewise, some challenges for IDR are of a social and not intellectual nature. For example, the discipline-based department system at most universities often incentivizes disciplinary research over IDR. These social dimensions may be less relevant to an A-IDR system. AI may be more skilled at IDR than humans because AI lacks social reasons to cluster into disciplines. However, if an A-IDR system derives its knowledge from the corpus of human scholarship, then it may learn and perpetuate disciplinary divisions, just as current AI systems learn and perpetuate human biases obtained from other human-produced datasets.

The remainder of this section surveys some cognitive challenges of IDR. All forms of research can be cognitively difficult, but IDR poses some distinct challenges that make it especially challenging for human researchers. Note that these challenges also apply to transdisciplinary research and to a lesser degree to multidisciplinary research. For more general discussion of interdisciplinarity, multidisciplinarity, and transdisciplinarity, see e.g., Hoffmann et al. (2013), McGregor (2014), Bernstein (2015), Lawrence (2015), Scholz and Steiner (2015), Menken and Keestra (2016), and Frodeman (2017); for an alternative and more critical perspective, see Jacobs (2013).

2.1 Disciplinary Divides

Different academic disciplines commonly approach the same topic from different perspectives, using different conceptual paradigms and different language, featuring different intellectual traditions and standards, and often covering different aspects of the same topic. This creates the cognitive challenge of integrating the disparate input into a unified understanding of a topic (Bracken and Oughton 2006; Keestra 2017; MacLeod 2018). As a consequence, IDR requires a laborious and often frustrating process of translation between different disciplines. It is a translation of language as well as of intellectual norms and epistemic perspectives.

Much of the challenge comes from the fact that human expertise builds from years of focused study and practice. The human mind can achieve high levels of performance on specific tasks by forming complex representations of the task and its surrounding context—for example, chess masters "chunking" (Chase and Simon 1973) together complex patterns of chess pieces. The sheer difficulty of developing expertise precludes individual researchers from being expert on the full range of subjects that are relevant to complex societal problems. Furthermore, the very nature of human expertise can make it harder for experts in one discipline to understand other disciplines. Indeed, the term "discipline" implies a certain disciplining of the human mind to think in certain ways. A human mind disciplined to think in one way can struggle to think in another way.

Compounding the problem is the fact that academia is traditionally divided into thematic disciplines that do not map to societal problems. A scholar is typically trained and employed as, for example, a computer scientist or a social scientist, not as an expert on the social implications of computers. Academia has made some effort to restructure toward IDR, but the traditional disciplines still dominate. As a consequence, disciplinary divides are even greater than they in principle need to be.

2.2 Massive Literatures

Even if a researcher or research team is able to parse the disparate disciplinary contributions to a particular topic, there remains the challenge of reading the relevant literature. Even within a single discipline, literature review can be a highly laborious task. For complex interdisciplinary topics, it quickly becomes overwhelming.

For example, the Intergovernmental Panel on Climate Change (IPCC) is a body of top global warming researchers from around the world. It produces periodic syntheses of the global warming literature for high-level policy audiences. IPCC reviews are massive exercises, involving hundreds of scientists each putting in hundreds of hours over periods of up to 5 years (Victor 2015). Despite this massive scale, the literature is now at a point where the IPCC's teams of researchers are struggling to keep up (Minx et al. 2017). Climate change is an especially complex issue with a large literature even by IDR standards, but the IPCC is also an unusually large IDR project. Other IDR projects with narrower scopes and smaller teams face similar literature challenges.

The IPCC is distinctive in its effort to synthesize the entire peer-reviewed literature on a complex interdisciplinary topic. Researchers who merely wish to contribute to the literature do not need to read so much. However, gaining a basic understanding of global warming in all its facets still requires reading at least some literature across numerous subjects in natural science, social science, engineering, the humanities, policy, etc. This is an extensive undertaking, and it is compounded by the added challenge of translating across the disciplinary divides.

2.3 Peer Review

IDR poses distinct peer-review challenges. IDR commonly works at the interface of multiple traditional academic disciplines. In many cases, it is uncommon to have expertise in each of the disciplines. Indeed, the person or team performing the IDR may be the only one in the world working across that particular set of topics. For example, this paper works at the intersection of AI and IDR; I am not aware of any other research that does this.

The eclectic mixes of topics found in IDR make it unusually difficult to peer review (Laudel 2006; Pautasso and Pautasso 2010; Holbrook 2017). If no other researchers have expertise across the range of topics covered in the work, then it may be necessary to form an interdisciplinary team of reviewers, but then this team must overcome its own disciplinary divides, which is a substantial task just to complete a single review. Or, the task of working across the disciplinary divides may fall to the people handling the submissions—journal editors, grant program managers, etc.—but then this just adds to their burden.

It has been suggested that the difficulty of reviewing IDR, combined with the tendency of researchers to favor work from their own specialty, biases reviewers against IDR (Laudel 2006). One study has found that IDR grant proposals are indeed funded at a lower rate (Bromham et al. 2016). This is another way in which the current academy may be systematically biased against IDR, limiting its contribution to addressing important societal problems.

2.4 Transfer

Some research aims to transfer insights from the study of one societal issue to another. Transfer is seldom discussed in literature about IDR, which tends to focus on one issue at a time (but see Krohn 2017). Transfer is not a panacea, especially given the distinct attributes that each societal issue has, which limits the extent to which insights transfer from one issue to another. Nonetheless, it can be a powerful way to study societal issues. Transfer can be especially helpful for newly emerging issues that have not yet built up a robust literature, such as issues involving emerging technologies. For example, Altmann and Sauer (2017) transfer insights from strategic studies of nuclear weapons to the study of newly emerging autonomous weapon systems. Baum (2017a) transfers insights from the psychology of promoting environmentally beneficial behaviors among the lay public to the study of how to encourage AI researchers to pursue socially beneficial AI designs.

The psychology of transfer finds that people tend to be more successful at it when there is relatively little "distance" between the domain transferred *from* and the domain transferred *to* (Perkins and Salomon 1992). For example, it may be easier to transfer insight from one weapon system to another (as in Altmann and Sauer 2017) than it is to transfer insight from lay public environmental behavior to expert behavior on technology development (as in Baum 2017a). Unfortunately, much of the insight available to be transferred lies at a considerable conceptual distance. A major challenge is simply recognizing the cross-issue similarity. Indeed, another finding from the psychology of transfer is that people often struggle to apply lessons to "problem isomorphs", i.e., to another domain that is functionally equivalent but different in appearance (Simon and Hayes 1976). People even struggle with problem isomorphs when both domains are presented to them. IDR transfer is more difficult because researchers typically must identify the two domains for themselves. On top of that, IDR transfer can require bridging three sets of epistemic divides: the divide between the two issues and the divides within each of the two issues. This can make for an especially difficult cognitive task.

3. Near-Term A-ID

This section discusses A-ID that currently exists or could exist with relatively straightforward extensions of current technology. Note that this section takes a relatively broad view of what qualifies as AI. By narrower standards, some of what is presented here may not qualify as AI. The exact definition and scope of AI is complex and controversial and beyond the scope of this paper; for further discussion see, e.g., Legg and Hutter (2007) on definitions and McCorduck (2004) on how the scope of what is considered to classify as AI has gotten narrower over time.

3.1 Search Engines

Perhaps the most ubiquitous form of A-ID is the search engine for internet and scholarly databases, including Dimensions.ai, Euretos, Google Scholar, IRIS.AI, Microsoft Academic, Omnity, Semantic Scholar, SourceData, and Web of Science. Search engines are valuable for many forms of research, but they are especially valuable for IDR. In narrow disciplinary fields, researchers need to follow a relatively small body of literature. They may even be able to keep up with the relevant literature without the use of search engines, instead identifying literature by reading certain journals and sharing relevant literature within peer communities. In contrast, IDR frequently pushes researchers outside their areas of familiarity and requires them to identify relevant literature from within extremely large bodies of work.

Search engines can help with each of the IDR challenges described in Section 2. They can provide some help to the challenge of overcoming disciplinary divides by providing a tool to explore the literatures of other disciplines, though this leaves open the challenge of understanding publications in unfamiliar disciplines. They can be used to navigate the massive literatures on interdisciplinary topics, such as by searching for multiple keywords to find literature on specific themes for a societal problem—for example, "global warming" and "cost-benefit analysis". They can be used to identify peer reviewers—for example, the authors of papers on global warming and cost-benefit analysis. And, with some creativity, they can support the transfer of insights across societal problem—for example, one could search for literature on global warming and cost-benefit analysis to identify insights that can be transferred to cost-benefit analysis of other global environmental issues.

Search engines could be improved. To my eyes, an important area for improvement is in the handling of synonyms and, more generally, the identification of terminology. When searching for literature on an unfamiliar topic, identifying the right keywords to search for can be a major impediment. Once one knows the right keywords, relevant literature often flows abundantly. It would be especially valuable to have tools that could identify keyword synonyms used by other disciplines. Handcrafted resources, such as Wikipedia, can be helpful, but such resources are not available for all keywords. Automated tools for this could be quite helpful.

Synonym identification is an active subject of AI research. One approach is to analyze patterns in the text surrounding a word, known as "distributional word vectors" or "word embeddings", on grounds that synonyms can be used in the same way and therefore tend to be surrounded by similar text (Leeuwenberg et al. 2016; Mohammed 2020). However, this work uses large linguistic datasets (i.e., large collections of text). Much less text is available for identifying synonyms in academic literature, especially for the many specialized concepts and small subfields. Indeed, the rarity of specialist terms can be a distinguishing characteristic of academic publications. One academic search engine, Omnity (https://www.omnity.io), uses the rarest words of a document to identify other documents with a similar mix of rare words on grounds that documents with the same rare words are likely to be related (Perkel 2017). Omnity's approach may work well within a discipline, but it may be less well suited for search across disciplines that use different terms for the same concepts. Further research may be needed to develop techniques for synonym identification in small academic literatures.

3.2 Recommendation Engines

At least two forms of recommendation engines are currently available for research. The first provides recommendations of publications. For example, Google Scholar provides publication recommendations that are customized for individual user profiles, and Elsevier provides publication recommendations that are customized for specific publications on its ScienceDirect website. A more specialized example is the project http://x-risk.net that provides recommendations of publications in the field of catastrophic risk (Shackelford et al. 2020). These recommendations provide an additional way for researchers to identify relevant literature on the topics they are studying. Note that the recommendations cover topics that researchers have already identified, via their user profiles or the publications they are already looking at, so they are less valuable for the IDR challenge of exploring unfamiliar lines of research. Therefore, in terms of the challenges described in Section 2, these recommendation engines can be especially valuable for handling massive literatures, and they may be of relatively limited value for overcoming disciplinary divides.

A second form of research recommendation engine recommends potential peer reviewers to journal editors. This is used, for example, in the Elsevier Evise system, which was launched in 2015. Insofar as Evise provides good recommendations of peer reviewers, it can lighten the burden on journal editors, as noted, for example, by the editor of the interdisciplinary Elsevier journal *Ecological Economics* (Hukkinen 2017). Another example is the Toronto Paper Matching System developed for matching papers to reviewers for computer science conferences (Charlin and Zemel 2013). While the Toronto Paper Matching System is used predominantly for computer science, it potentially could be expanded to more IDR domains. The value of recommendation engines for IDR could be especially large due to the difficulty of identifying appropriate reviewers for IDR.

One way to improve interdisciplinary recommendation engines would be to streamline the process of building specialized recommendation engines. The recommendation engine presented by Shackelford et al. (2020) uses a custom artificial neural network trained with a hand-coded dataset of articles found to be relevant to the field of catastrophic risk by an open ("crowdsourced") team of researchers in the field. Other fields could apply the same approach. Hand-coding the training dataset is laborious but is vital for aligning recommendation engine results with the interests of people in the field. The labor required for this may be large, but it does not require any skills other than knowledge of the field. In contrast, developing the artificial neural network does require more specialized skills. Fields that lack people with these skills may struggle to develop their own recommendation engines. Therefore, the process of building field-specific recommendation engines could be streamlined by developing an artificial neural network that could be trained with data from any field and in particular by developing a more accessible user interface for it.

3.3 Automated Content Analysis

A final application of AI to research is the use of machine learning for analyzing the content of the literature on a given topic, known as "automated content analysis" (Nunez-Mir et al. 2017). Essentially, it treats academic literature as an instance of "big data", sometimes referred to as "big literature" (Nunez-Mir et al. 2016). Automated content analysis analyzes clusters of key words and phrases to produce statistical trends in literatures. This can be of value for learning what a literature has tended to find, where gaps may lie, etc. It can also be used for identifying relevant literature on a topic and in this capacity may be considered a form of search engine. It is of particular interest in the context of systematic reviews, in which the aim is to provide a complete and unbiased account of the literature on a particular topic. In terms of the challenges described in Section 2, this can be of high value for handling the massive literatures of IDR.

While automated content analysis has been used mainly for narrow disciplinary research, it has been used for IDR studies of forest ecology (Nunez-Mir et al. 2017) and participatory design (Tuhkala et al. 2018). The Nunez-Mir et al. (2017) study is illustrative of the possibilities. It analyzes the text of 14,855 abstracts in 7 forestry journals. It uses a mix of supervised and unsupervised seeding, meaning that central concepts are obtained via a mix of human input and algorithmic text mining. Subsequent algorithmic analysis assesses the preponderance of the concepts in the abstract database. A primary finding of the study is that most of the abstracts do not show an interdisciplinary approach and very few considered social dimensions of forest ecology. In consideration of this finding, Nunez-Mir et al. (2017) call for more IDR on forestry, especially its social dimensions.

Note that the Nunez-Mir et al. (2017) study does not reveal anything about the nature of forests or their interdisciplinary attributes beyond the list of central concepts. The output is a map of the literature, not an interpretation of what the literature means and what its societal implications are. This is a general limitation of current automated content analysis. Thus, as Sutherland and Wordley (2018, p.366) write:

> Advances in artificial intelligence and machine learning could make it easier to perform tasks such as locating papers for defined topics using search terms, categorizing papers as relevant for further consideration, and producing systematic maps. But for all fields, assessing the quality of individual studies, writing up summaries and so on will continue to require skilled humans, at least for the foreseeable future.

The above remark about "the foreseeable future" suggests that the AI needed for interpreting literature and its societal implications is significantly beyond the capacity of current AI. My own judgment is that this is probably correct. Therefore, interpretation is considered in the following section on intermediate-term A-ID.

4. Intermediate-Term A-ID

Intermediate-term A-ID systems would go significantly beyond current capabilities, but not so far beyond that they would constitute "strong" A-ID with human-level or super-human-level capabilities. Intermediate-term is of interest because it could be of greater value to IDR than near-term A-ID while avoiding the downsides associated with strong AI (see Section 5). Ideally, designs for intermediate-term A-ID would derive at least in part from existing AI techniques such that current AI researchers could work toward their development. There can also be value to envisioning intermediate-term A-ID that requires new techniques but would stop short of strong AI. Of course, the sooner new A-ID capabilities become available, the sooner they can contribute to addressing important societal problems.

As stated in the introduction, I am not a computer scientist, and so I am less confident in my thoughts on the design of intermediate-term A-ID systems. Nonetheless, a few basic thoughts on computer science factors are offered as a contribution to the topic, alongside some social factors that I can comment on more confidently.

4.1 Interpretation

One direction to extend current A-ID capabilities would be via the interpretation of research publications. As Sutherland and Wordley (2018) explain, this is a task currently left for human researchers (see Section 3.3). However, progress in the field of machine reading (a.k.a. natural-language understanding/interpretation) could change this. For example, one project to automate interpretation is the Elsevier-sponsored ScienceIE (Augenstein et al. 2017). The current focus of ScienceIE is on identifying phrases and relations between phrases within narrow disciplinary literatures. Its aims are ambitious: "Say you have a question about a paper: A machine learning model reads the paper and answers your question" (Augenstein as interviewed by Stockton 2017).

Automated interpretation of research publications would be of considerable value for many forms of research, but it may be of particularly large value for IDR, due to the large and diverse literatures involved in IDR. An integrated system for automated content analysis and interpretation would be especially valuable for synthesizing the insights contained in vast and diverse IDR literatures. Such capability could drastically reduce the burden of human researchers in IDR synthesis projects like the IPCC and could likewise enable similar synthesis projects across a wider range of societal problems. In terms of the IDR challenges described in Section 2, interpretation would be especially valuable for handling disciplinary divides and massive literatures, especially if combined with automated content analysis.

Publications can be interpreted in many different ways, some of which may be easier for A-IDR. For example, interpreting whether a publication has presented results that are statistically significant may be easier for A-IDR than interpreting whether the publication has presented results that could help address some societal problem. It may be the case that current AI paradigms based on neural networks will struggle with the interpretation of complex interdisciplinary texts because, as discussed by Marcus (2018), neural networks are skilled at finding statistical relationships in complex datasets but struggle to handle causal relationships, hierarchies, and open-ended environments, all of which are important attributes of interdisciplinary texts. However, the extent to which this is indeed the case is left as a question for future research on A-ID.

4.2 Translation

Given the difficulty of translating across the epistemic divides that exist between academic disciplines (Section 2.1), it could be of considerable value to have A-ID systems that could assist with the translation. A-ID translation would convert between the jargons of different disciplines or between that of one jargon and a more widely accessible plain English. This would reduce the linguistic burden of IDR, thereby facilitating researchers to work across a wider range of disciplines, either alone or in teams. It would also be of value for peer reviewers tasked with reviewing interdisciplinary work that includes topics outside the reviewers' own areas of expertise.

Potentially, A-ID translation could build off existing work on machine translation between languages (e.g., English to Spanish). An active area of research is for unsupervised machine translation (e.g., Lample et al. 2018; summarized in plain English by Ranzato et al. 2018), in which translation does not require large datasets of pre-translated text. Unsupervised machine translation would be important for interdisciplinary translation because large datasets generally do not exist and would be expensive to produce.

A-ID translation may require techniques that go beyond cross-language machine translation because the text of different disciplines commonly varies by conceptual basis and not just by language. In other words, much of the challenge of reading text from other disciplines is that readers do not know the concepts that the words are describing or at least are not accustomed to thinking in terms of these concepts. Interdisciplinary translation is about explaining unfamiliar concepts just as much as it is about explaining unfamiliar terminology. In contrast, cross-language machine translation generally involves concepts that are familiar in both languages. For example, Ranzato et al. (2018) presents the translation of the phrase "cats are lazy" from English to Urdu; presumably Urdu speakers are already familiar with the concept of a lazy cat. Therefore, successful A-ID translation may need to include a capacity for concept identification and translation that goes substantially beyond cross-language machine translation.

4.3 Transfer

The challenge of interdisciplinary transfer (Section 2.4) maps neatly to the AI topic of transfer learning (Pan and Yang 2009). Both involve the same essential process of applying knowledge gained in one domain to another domain with similar features, especially to avoid the resource-intensive process of building complex knowledge sets for multiple domains. Complex societal problems often have some similarities with each other. Human researchers can barely scratch the surface of the potential of interdisciplinary transfer due to its cognitive difficulty and the very large number of combinations of similar societal problems. If AI transfer learning could be applied to IDR, it could open up vast possibilities for learning about societal problems.

Current AI transfer learning involves tasks that are rather far removed from IDR. For example, one common AI transfer learning task is image recognition (Zamir et al. 2018). Whereas images are readily expressed in statistical terms (e.g., via the binary representation of a digital image file), societal problems are not. To be sure, there are ways to express societal problems in some quantitative terms, such as via cost-benefit analysis for the evaluation of solutions. But much of the insight involved in interdisciplinary transfer is qualitative. A-ID transfer learning would need reliable ways to quantify this insight. It would also need some degree of translation, given the differing terminology and other linguistic constructs that can exist across societal problems.

There is some reason to believe that A-ID transfer learning could be quite difficult to achieve. Computers are very capable at churning through large numbers of combinations of predefined concepts, just as they are very capable at churning through large numbers in general. However, they may struggle at identifying the sorts of connections between concepts that humans find meaningful. In a discussion of creativity in humans and computers, Boden (2009) argues that humans are relatively skilled, and computers relatively unskilled, at "combinatorial creativity", meaning creativity in making new associations between seemingly unrelated concepts. Humans have rich mental models of the world built up throughout lifetimes of experience and observation, which enable us to devise analogies and imagery and other forms of combinatorial creativity. Endowing computers with anything similar is very difficult, and computers likewise struggle at combinatorial creativity. Interdisciplinary transfer is, at least in part, a form of combinatorial creativity, and so A-ID systems may struggle with it.

5. Long-Term A-ID

As an abstract matter, it is not hard to imagine some future AI that is capable of performing the full range of cognitive tasks involved in IDR. A strong AI/human-level AGI would presumably be able to do IDR at least as well as humans could. Whereas "narrow" AI is only intelligent for a narrow range of cognitive tasks, AGI is general in the sense of being able to perform a wide range of cognitive tasks. Human-level AGI could perform the same breadth of cognitive tasks as human minds and with the same skill as human minds. Such an AI could do any other cognitive task that humans can do, so it would presumably also be capable of IDR, potentially even so capable as to effectively solve all of society's problems. The AI may not be operating in complete isolation from humans—studies of human problems would likely require some human participation—but the AI may be able to play the entire suite of cognitive roles currently played by human interdisciplinary researchers.

The prospect of solving societal problems is one common motivation for developing AGI. This is seen in the stated goals of active AGI research and development projects. For example, the AGI project AIDEUS observes that the "limitation of intellectual possibilities is fully experienced by scientists" and that this "concerns all problems of people – from traffic jams to economic crises and wars". That is very much in the spirit of the present paper. AIDEUS states this as a motivation for building strong AI.[1] As another example, DeepMind envisions AI for "helping humanity tackle some of its greatest challenges, from climate change to delivering advanced healthcare".[2] Finally, Jeff Hawkins of Numenta calls for AI to help humanity "face challenges related to disease, climate, and energy" (Hawkins 2017). Note that these and other examples can be obtained from Baum (2017b).

As valuable as AGI could be for IDR and for addressing complex societal problems, it faces several important downsides. First is the technical difficulty of building AGI. Despite the numerous projects working on AGI, the current state of the art remains quite far removed from it. It is not clear if or when AGI would be built; projections of decades or longer are common (Baum et al. 2011; Grace et al. 2018). That is a long time to wait for solutions to major societal problems. For some problems, such as nuclear war or extreme pandemics, the effects could be so severe that civilization would fail before it has the chance to build AGI.

The desire to use AGI for societal problems could be additionally problematic by putting a dangerous time pressure on AGI development. Herein lies a dilemma. On one hand, earlier AGI

[1] http://aideus.com/community/community.html
[2] https://deepmind.com/blog/learning-through-human-feedback

would help with other societal problems. On the other hand, a rushed AGI could skimp on built-in safety measures and ethical design, increasing the risk of adverse or even catastrophic harm caused by the AGI. To the extent that major societal problems can be addressed without AGI, this buys AGI developers more time to get it done right.

As an aside, note that some scholarship on IDR may be of value for the safe development of AGI, noting that the project of building AGI is itself an interdisciplinary task. For example, Keestra (2017) documents how a failure of interdisciplinary communication contributed to the 1986 space shuttle Challenger disaster and analyzes how future interdisciplinary communication could go better. Communication among those developing and launching AGI may benefit from this analysis.

Careful AGI design may be needed even for AGI systems that exist only to provide input on societal problems. This is a theme of the literature on the prospect of superintelligent "oracle" AIs that are designed to only answer questions that humans pose to it (Armstrong et al. 2012; Yampolskiy 2012). The oracle design is sometimes conceived as a way to make AGI safe. The hope is that by restricting the AI to only answering questions, humanity can prevent the AI from taking dangerous physical actions to alter the world. However, studies of superintelligent oracles have raised the concern that they could still be dangerous, for example, by manipulating the people who interact with it (*ibid.*). The prospect of strong A-ID being dangerous in this way is another reason to explore near- and intermediate-term A-ID that can help address societal problems without raising the concerns associated with strong AI.

Finally, it should be noted that the development of AGI is itself an interdisciplinary issue that would benefit from improved capacity for IDR. The development of AGI involves computer science, ethics, risk management, and more. Important questions include how best to design the AGI, how best to manage the human teams that are working on the design, and when to launch the AGI given the potential benefits and harms that could come from it. Insofar as near- and intermediate-term A-ID could improve the available answers to these sorts of questions, it could improve the outcomes from AGI development. This gives another reason to pursue near- and intermediate-term A-ID, including the concepts outlined in this paper.

6. Discussion

The limitations of near-term A-ID and the difficulties and risks of long-term A-ID suggest an important role for intermediate-term A-ID. A-ID interpretation, translation, and transfer could offer powerful capabilities for IDR and, more generally, for understanding and addressing complex societal problems. They could combine with existing A-ID systems (search engines, recommendation engines, and automated content analysis) to provide support across all of the cognitive challenges of IDR. This would not necessarily automate the full suite of IDR tasks—it could still leave important roles for human researchers—but it could nonetheless be of considerable value to IDR.

On the other hand, some of these new capabilities could push uncomfortably toward AGI. Yampolskiy (2013) argues that the understanding of natural language is "AI-complete", meaning that any AI system that is capable of understanding natural language would necessarily also be an AGI with a wide range of other capabilities. Similarly, Armstrong et al. (2012, p.305) postulate that "the task of advancing any field involving human-centred issues—such as economics, marketing, politics, language understanding, or similar—is likely to be AI-complete", and Stockton (2017) suggests that an AI capable of peer-reviewing research papers would require AGI. Exactly what, if anything, is AI-complete may be controversial and is beyond the scope of this paper, but it is nonetheless worth considering the implications of these arguments.

These arguments imply that significant portions of A-ID interpretation, translation, and transfer may be impossible without AGI. Interpretation inherently involves understanding natural language, and so there may be very little A-ID interpretation that can be done without AGI. The translation of concepts may require understanding natural language, and so, without AGI, A-ID translation may be limited to the translation of terminology. Finally, transfer may also require understanding natural language and therefore AGI unless the societal problems of IDR can be reduced to a simpler quantitative form. The extent to which A-ID interpretation, translation, and transfer can be done without AGI is an important open question.

Another concern is that intermediate-term A-ID could introduce certain biases into the research. This could occur if the A-ID interprets or evaluates research differently than human interdisciplinary researchers. For example, Hukkinen (2017) expresses concern that AI may succeed at assessing the intellectual rigor of research publications but struggle at assessing their significance for societal problems. Indeed, the ScienceIE project is explicitly framed in terms of *scientific* publications, but the interpretation of publications in other (non-science) areas that are essential to IDR for societal problems (such as ethics and public policy) require different ways of thinking and potentially different AI techniques. An A-ID system that can perform well on scientific rigor but poorly on ethics and policy could lead to weak or even harmful insight on how to address societal problems. More generally, intermediate-term A-ID systems must be carefully evaluated to ensure that they are producing constructive assistance to IDR. This will likely benefit from close collaboration between computer science developers of A-ID and human interdisciplinary researchers and potentially also people with other backgrounds, including stakeholders from outside academia.

The inclusion of non-academic stakeholders is especially important with an eye toward actually addressing societal problems instead of just advancing the academic study of them. The existence of IDR literature does not on its own imply that the problems actually get addressed. Absent close engagement from non-academic stakeholders, there is a risk of producing solutions that work well in theory but not in practice. Additionally—and importantly—engagement with non-academic stakeholders can result in different formulations of what the problems are in the first place. For these reasons, IDR—and especially transdisciplinary research—often calls for stakeholder participation in the research process.

Readers should be aware that IDR often involves controversy, especially when it reaches out beyond the academy. Complex and important societal problems often involve heated disagreements between different stakeholders: between environmentalists and the fossil fuel industry over global warming, between rival geopolitical powers over nuclear weapons, etc. In some cases, there can be clever technical solutions that please everyone or at least do not upset anyone. However, more typically, solutions involve difficult tradeoffs between competing factions and fundamental moral values. Readers are encouraged to consider contributing to A-ID, but they should do so with an awareness of all that this entails.

In the extreme case, A-ID could even be misused so as to make societal problems worse. For example, some of the same IDR that can help environmentalists refine their messages can also do the same for the fossil fuel industry, a dynamic that may be compounded by the extensive financial resources that the industry has at its disposal. Existing interdisciplinary research communities have often handled this by engaging with the ethical dimensions of societal problems and striving for ethically sound IDR practice. Work on A-ID should do the same.

Finally, it should be noted that A-ID could help with all IDR, not just IDR on societal problems. As discussed in Section 2, IDR is perhaps most commonly performed for societal problems, but some IDR is oriented toward scientific and intellectual progress without any particular regard for societal problems. Such A-ID may be significant from the perspective of philosophy of science; this matter is left for future research.

7. Conclusion

This paper has articulated the concept of A-ID, outlined its importance for helping human researches cope with and overcome the considerable cognitive challenges of IDR, and explained how this can help address complex societal problems. A-ID exists today in the form of search engines, recommendation engines, and automated content analysis, but this leaves human researchers with the bulk of the work. Insofar as A-ID systems could contribute more to IDR, this could be a major contribution to addressing major societal problems. AI researchers interested in applying AI to societal problems may find this a worthy area of application. This paper describes some potential A-ID research directions; readers with more background in computer science than myself may have further suggestions. However, there are some difficult issues involved in A-ID that researchers should be cautious about, especially regarding the controversies inherent to complex societal issues, the potential for misuse of A-ID, and the potential relation of A-ID to strong AI/AGI. For their part, interdisciplinary researchers can contribute by using A-ID tools in their research and by contributing to efforts to improve the tools. Advancing A-ID is itself an interdisciplinary task that will benefit from contributions from people with a variety of backgrounds.

Essential IDR cognitive tasks may be AI-complete, meaning that A-ID capable of these tasks would need to be strong AI. These tasks could include the interpretation of research publications, the translation of research concepts across disciplinary divides, and the transfer of IDR insights to new societal problems. If weak/narrow A-ID is unable to handle these tasks, or other essential IDR tasks, then human interdisciplinary researchers may be largely on their own until the advent of strong AI. This would mean that, despite all the disparate ongoing accomplishments of narrow AI, it would make relatively little contribution to addressing society's most complex and important problems. Alternatively, if there could be progress on A-ID that would not require AGI, then this could be a major contribution and would be a worthy focus for the field of AI.

Acknowledgments

Robert de Neufville, Roman Yampolskiy, Stuart Armstrong, Daniel Filan, Gorm Shackelford, Machiel Keestra, Mahendra Prasad, Josh Cowls, and two anonymous reviewers provided helpful comments on an earlier version of this paper. Any remaining errors are the author's alone.

References

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. Survival, 59(5), 117-142.

Armstrong, S., Sandberg, A., & Bostrom, N. (2012). Thinking inside the box: Controlling and using an oracle AI. Minds and Machines, 22(4), 299-324.

Augenstein, I., Das, M., Riedel, S., Vikraman, L., & McCallum, A. (2017). Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. Proceedings of the International Workshop on Semantic Evaluation (SemEval at ACL 2017). https://arxiv.org/abs/1704.02853.

Baum, S.D. (2017a). On the promotion of safe and socially beneficial artificial intelligence. AI & Society, 32(4), 543-551.

Baum, S.D. (2017b). A survey of artificial general intelligence projects for ethics, risk, and policy. Global Catastrophic Risk Institute Working Paper 17-1.

Baum, S. D. (2018). Reconciliation between factions focused on near-term and long-term artificial intelligence. AI & Society, 33(4), 565-572.

Baum, S.D. (2020). Medium-term artificial intelligence and society. Information, 11(6), 290, DOI 10.3390/info11060290

Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. Technological Forecasting and Social Change, 78(1), 185-195.

Bernstein, J. H. (2015). Transdisciplinarity: A review of its origins, development, and current issues. Journal of Research Practice 11(1), article R1.

Boden, M. A. (2009). Computer models of creativity. AI Magazine, 30(3), 23-34.

Bracken, L. J., & Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. Transactions of the Institute of British Geographers, 31(3), 371-382.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. Nature, 534, 684-687.

Cave, S., Ó hÉigeartaigh, S. S. (2019). Bridging near- and long-term concerns about AI. Nature Machine Learning, 1(1), 5-6.

Charlin, L., & Zemel, R. S. (2013). The Toronto Paper Matching System: An automated paper-reviewer assignment system. International Conference on Machine Learning (ICML) 2013, Workshop on Peer Reviewing and Publishing Models.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive psychology, 4(1), 55-81.

Committee on Facilitating Interdisciplinary Research and Committee on Science, Engineering, and Public Policy (2005). Facilitating Interdisciplinary Research. Washington, D.C.: National Academies Press.

Crease, R. P. (2017). Physical sciences. In R. Frodeman (Ed.) The Oxford Handbook of Interdisciplinarity: Second Edition (pp. 71-87). Oxford University Press, Oxford.

Frodeman, R. (Ed.) (2017). The Oxford Handbook of Interdisciplinarity: Second Edition. Oxford University Press, Oxford.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729-754.

Hawkins, J. (2017). What intelligent machines need to learn from the neocortex. IEEE Spectrum, 2 June. https://spectrum.ieee.org/computing/software/what-intelligent-machines-need-to-learn-from-the-neocortex.

Hoffmann, M. H., Schmidt, J. C., & Nersessian, N. J. (2013). Philosophy of and as interdisciplinarity. Synthese, 190(11), 1857-1864.

Holbrook, J. B. (2017). Peer review, interdisciplinarity, and serendipity. In R. Frodeman (Ed.) The Oxford Handbook of Interdisciplinarity: Second Edition (pp. 485-497). Oxford University Press, Oxford.

Hukkinen, J. I. (2017). Peer review has its shortcomings, but AI is a risky fix. Wired, 30 January, https://www.wired.com/2017/01/peer-review-shortcomings-ai-risky-fix.

Jacobs, J. A. (2013). In defense of disciplines: Interdisciplinarity and specialization in the research university. Chicago: University of Chicago Press.

Keestra, M. (2017). Metacognition and reflection by interdisciplinary experts: Insights from cognitive science and philosophy. Issues in Interdisciplinary Studies 35, 121-169.

Klein, J. T. (2017). Typologies of interdisciplinarity: The boundary work of definition. In R. Frodeman (Ed.) The Oxford Handbook of Interdisciplinarity: Second Edition (pp. 21-34). Oxford University Press, Oxford.

Krohn, W. (2017). Interdisciplinary cases and disciplinary knowledge: Epistemic challenges of interdisciplinary research. In R. Frodeman (Ed.) The Oxford Handbook of Interdisciplinarity: Second Edition (pp. 40-52). Oxford University Press, Oxford.

Lample, G., Ott, M., Conneau, A., Denoyer, L., & Ranzato, M. A. (2018). Phrase-based & neural unsupervised machine translation. https://arxiv.org/abs/1804.07755

Laudel, G. (2006). Conclave in the Tower of Babel: how peers review interdisciplinary research proposals. Research Evaluation, 15(1), 57-68.

Lawrence, R. J. (2015). Advances in transdisciplinarity: Epistemologies, methodologies and processes. Futures 65 (2015) 1-9.

Leeuwenberg, A., Vela, M., Dehdari, J., & van Genabith J. (2016). A minimally supervised approach for synonym extraction with word embeddings. Prague Bulletin of Mathematical Linguistics 105, 111-142.

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds & Machines, 17(4), 391-444.

MacLeod, M. (2018). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese, 195(2), 697-720.

Marcus, G. (2018). Deep learning: A critical appraisal. https://arxiv.org/abs/1801.00631

McCorduck, P. (2004). Machines Who Think: 25th Anniversary Edition. Natick, MA: AK Peters.

McGregor, S. L. T. (2014). Introduction to special issue on transdisciplinarity. World Futures, 70(3-4), 161-163.

Menken, S., & Keestra, M. (Eds.) (2016). An Introduction to Interdisciplinary Research: Theory and Practice. Amsterdam: Amsterdam University Press.

Minx, J. C., Callaghan, M., Lamb, W. F., Garard, J., & Edenhofer, O. (2017). Learning about climate change solutions in the IPCC and beyond. Environmental Science & Policy, 77, 252-259.

Mohammed, N. (2020). Extracting word synonyms from text using neural approaches. International Arab Journal of Information Technology, 17(1), 45-51

Newell, W. H. (2001). A theory of interdisciplinary studies. Issues in Integrative Studies, 19, 1-25.

Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N., & Fei, S. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. Methods in Ecology and Evolution, 7(11), 1262-1272.

Nunez-Mir, G. C., Desprez, J. M., Iannone III, B. V., Clark, T. L., & Fei, S. (2017). An automated content analysis of forestry research: are socioecological challenges being addressed? Journal of Forestry, 115(1), 1-9.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

Parson, E., Re, R., Solow-Niederman, A., & Zeide, E. (2019a). Artificial intelligence in strategic context: An introduction. AI Pulse, 8 February, https://aipulse.org/artificial-intelligence-in-strategic-context-an-introduction.

Parson, E., Fyshe, A., Lizotte, D. (2019b). Artificial intelligence's societal impacts, governance, and ethics: Introduction to the 2019 Summer Institute on AI and Society and its rapid outputs. AI Pulse, 26 September, https://aipulse.org/artificial-intelligences-societal-impacts-governance-and-ethics-introduction-to-the-2019-summer-institute-on-ai-and-society-and-its-rapid-outputs.

Pautasso, M., & Pautasso, C. (2010). Peer reviewing interdisciplinary papers. European Review, 18(2), 227-237.

Perkel, J. (2017). Omnity opens multilingual semantic searches up to academia. Nature Jobs, 12 January, http://blogs.nature.com/naturejobs/2017/01/12/omnity-opens-multilingual-semantic-searches-up-to-academia.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. International Encyclopedia of Education, Oxford, Pergamon Press, pp. 6452-6457.

Prunkl, C., & Whittlestone, J. (2020). Beyond near- and long-term: Towards a clearer account of research priorities in AI ethics and society. In Proceedings of the Third AAAI / ACM Annual Conference on AI, Ethics, and Society, New York.

Ranzato, M., Lample, G., Ott M. (2018). Unsupervised machine translation: A novel approach to provide fast, accurate translations for more languages. Facebook Code, 31 August, https://code.fb.com/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages.

Scholz, R. W., & Steiner, G. (2015). Transdisciplinarity at the crossroads. Sustainability Science, 10(4), 521-526.

Shackelford, G. E., Kemp, L., Rhodes, C., Sundaram, L., ÓhÉigeartaigh, S. S., Beard, S., Belfield, H., Weitzdörfer, J., Avin, S., Sørebø, D., Jones, E. M., Hume, J. B., Price, D., Pyle, D., Hurt, D., Stone, T., Watkins, H., Collas, L., Cade, B. C., Johnson, T. F., Freitas-Groff, Z., Denkenberger, D., Levot, M., Sutherland, W. J. (2020). Accumulating evidence using crowdsourcing and machine learning: A living bibliography about existential risk and global catastrophic risk. Futures, 116, 102508, DOI 10.1016/j.futures.2019.102508.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. Cognitive Psychology, 8(2), 165-190.

Stockton, N. (2017). If AI can fix peer review in science, AI can do anything. Wired, 21 February. https://www.wired.com/2017/02/ai-can-solve-peer-review-ai-can-solve-anything.

Sutherland, W. J., & Wordley, C. F. (2018). A fresh approach to evidence synthesis. Nature, 558, 364-366.

Tuhkala, A., Kärkkäinen, T., & Nieminen, P. (2018). Semi-automatic literature mapping of participatory design studies 2006-2016. In Proceedings of Participatory Design Conference (PDC'18), DOI 10.1145/3210604.3210621.

Victor, D. (2015). Climate change: Embed the social sciences in climate policy. Nature, 520(7545), 27-29

Wilson, E.O. (1998). Consilience: The Unity of Knowledge. New York: Knopf.

Yampolskiy, R. V. (2012). Leakproofing Singularity: Artificial Intelligence Confinement Problem. Journal of Consciousness Studies 19(1-2): 194–214

Yampolskiy, R. V. (2013). Turing test as a defining feature of AI-completeness. In X.-S. Yang (Ed.), Artificial intelligence, evolutionary computing and metaheuristics (pp. 3-17). Berlin: Springer.

Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3712-3722).