Will the real AI researcher please stand up?
Fields, Networks and Systems to Measure the Impact of Research Investments•

Jason Owen-Smith
University of Michigan
Institute for Research on Innovation & Science (IRIS)
jdos@umich.edu

• This work was supported by grants from the Alfred P. Sloan Foundation and the National Science Foundation (TF-233257, BCS-2318170). I build upon and quite shamelessly borrow from nearly 15 years of collaborative work with Julia Lane and Bruce Weinberg who both offered valuable comments on this piece. I also describe ongoing work on a recently funded project where Bruce and Anna Harvey are Co-PIs and Josh Hawley is a key collaborator. The work described here depends upon and cannot have been accomplished without the support and contributions of many people working at IRIS, Ohio State, NYU, and other sites where we have or have had productive collaborations including Johns Hopkins University and Elsevier. My name is on the document you are reading, but the good parts are the work of many hands. The errors and misunderstandings, however, are my own.

Executive Summary

Background.

- The CHIPS and Science Act (CSA) created high stakes measurement challenges about research investments' impacts on jobs, employers, education and workforce capabilities
- The core components needed to address CSA requirements exist and have demonstrated value individually for over a decade
- The Industries of Ideas project is building an end-to-end prototype system, piloting measures to meet CSA mandates and doing broad outreach to develop plans for scaling
- This measurement approach anchors technology and industry classification, measurement, and system design on the people research investments support
- We use scientific metadata to classify authors in specific fields, university administrative data (UMETRICS) to include all people paid by research grants, and state Unemployment Insurance (U-I) wage records to classify employers, industries and jobs
- Field and industry classifications that can work across CSA focal areas are essential but underdeveloped
- We propose a new strategy using social science concepts, network science measures, and deep learning models

Classification strategy

- Emerging technology areas are social and organizational research fields
- Research field boundaries and contents are defined based on people
- Classification has two-stages:
  - Domain experts help define a technology-specific "seed set" of researchers
  - a technology-agnostic semi-supervised deep learning model classifies additional investigators
- Model design and input data are based on a clear, explicable framework to ensure legibility for stakeholders including domain scientists, university and business leaders, state agencies and federal policy-makers
- Three dimensions of group membership - social proximity, affinity, and signaling – support a working data structure and model design

Use and Extensions

- Robust field classification will help the industries of ideas data system "follow-the-people" to classify industries associated with technologies, develop granular, timely measures to meet CSA requirements, and develop valuable products for universities, states and the nation.
- The system can expand to assess the impact of AI testbed and compute resources and systematize measures of AI model and software development
- Analytic extensions based on workforce data and deep learning models offer powerful new ways to examine "place-based" economic and industrial policy effects
- Strong but flexible governance mechanisms to facilitate responsible, secure data access and streamline collaboration will enable many currently unforeseen measurement and evaluation possibilities

Introduction: We are playing for high stakes.

Recent legislation that makes federal research investments in critical and emerging technologies a key lever for economic policy at massive scale has created pressing new measurement challenges. The CHIPS and Science Act (CSA) of 2022 authorized $81 billion for NSF, allocating about $1 in every $4 to the new Technology and Innovation Partnerships (TIP) directorate. TIP could become nearly 2.4 times bigger than today's largest NSF directorate (Mathematical and Physical Sciences), controlling nearly 60% of today's entire NSF research budget.

TIP's authorizing language foregrounds regional innovative capabilities, jobs, workforce and educational capacity, and broad concerns about equity and access. It articulates an ambitious goal, to " . . . advance research and development, technology development, and related solutions to address United States societal, national and geostrategic challenges, for the benefit of all Americans," and includes substantial measurement mandates. By 2027, CSA requires a National Academies review that, among other things, assesses:

- solutions to " . . . challenges with social, economic, health, scientific, and national security implications;"
- ". . . whether Federal investment in the key technology focus areas have resulted in new domestic manufacturing capacity and job creation;" and
- ". . . education and workforce development to support the key technology focus areas."

Yet CSA funding for TIP was an authorization, not an appropriation, leaving open what Congress ultimately funds. If Congress antes up, TIP represents an enormous bet on the economic and social impact of research investments. If not, the expectations CSA created may hollow out traditional NSF research areas in service to its goals, with potentially devastating consequences for the very societal, national, and geostrategic needs the act seeks to bolster. Either way, the new measurement and reporting requirements are here to stay. The ability to clearly, accurately, and reliably document what these investments do may help determine the act's lasting consequences. Today, to the best of my knowledge, nobody can do the work that is needed.

Key measurement challenges.

No one can do this yet because there are real challenges. The first is *classification*. The key technology focus areas are not scientific research fields as they are conceptualized in any current classification system. They also are not industries, at least not in any way we currently conceptualize them. Measuring inputs and outputs in a reliable fashion even for one field, AI, is difficult. We need to do it at scale, robustly, for all of them. Our current classifications need serious rethinking.

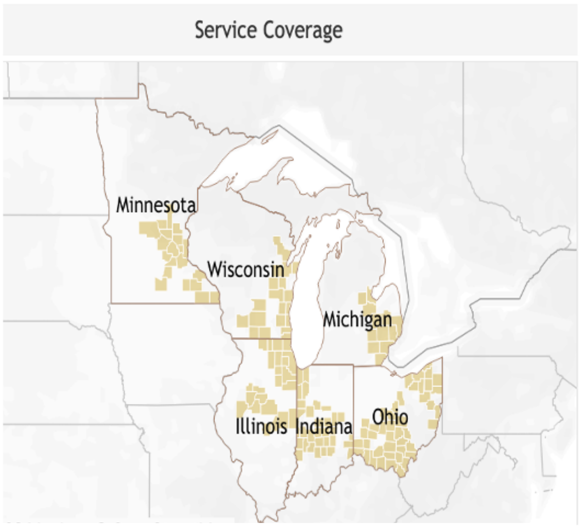

Figure 1. Example TIP RIE Finalist Region

The second challenge is about data requirements for *granularity* and *aggregation*. TIP's signature program is the Regional Innovation Engines (RIE). RIE applications currently define "regions" in terms of counties. They also specify a set of technologies. Figure 1 presents the "service area" for one RIE finalist proposal from TIP's website. "Great Lakes ReNEW" proposed to serve a "region" defined by clusters of counties that fall within and across the boundaries of six states in the upper Midwest.

Meeting CSA mandates for the RIE program requires the capability to identify: (1) the fields where relevant investments are made and innovative solutions produced as well as (2) the industries where jobs, manufacturing, education, and workforce capabilities might result. Moreover, that capacity must encompass any possible pairing of one or more CSA technologies with any combination of counties. Granular technology by county data must be aggregable to the national level to examine program effects. Finally, making a start on assessing impact necessitates, at a bare minimum, the ability to describe change over time.

Ideally, we would want causal estimates based on a variety of identification strategies. So a mature system should also include mechanisms for research access complete with means to ingest and link new data, appropriate privacy protections, essential data security, and all the other technical and legal infrastructure needed to allow responsible restricted data use. Which points to the third challenge.

This work will require *new institutional and governance mechanisms* that engage many autonomous, co-equal data owners – including states, corporations, federal agencies, and universities – in collaboratives with clear value propositions for each participant. Mechanisms to streamline formation and work by many different partnership configurations to meet different types of needs will be necessary.

These requirements mirror the regional multi-sector collaboration for national goals logic of the CSA. They also depart from traditional organizational models for national measurement, which typically center on the Federal Statistical Agencies.[1] One emerging, complementary alternative is a variety of data federalism characterized by regional networks of public-private partnerships anchored on relationships between universities and state agencies.[2,3,4] Such networks can very productively include multiple types of federal partners.

Conceptual framework for classification.

All the essential pieces of this system exist and have been individually proven through at least a decade of use. We are assembling the first full prototype that integrates all the core components to pilot measures that address CSA requirements. But a strong and clear classification strategy is essential. To do the job, that strategy must: (1) account for real differences among technologies; and (2) work across current CSA technologies and be expandable to new priority areas that arise. Three conceptual steps are necessary to meet that challenge.

Treat research fields as a form of social organization.

Technologies, like AI, are cases of a broader phenomenon long studied by social scientists. Such research areas are institutional fields: recognizable arenas for collaborative and competitive work done by diverse sets of people and organizations in and through evolving networks.[5] Understood as a form of social organization, all contemporary research fields have more commonalities than differences. Indeed, one the most influential early descriptions of such fields famously asked why we consistently observe so much similarity among social and economic competitors.[6]

Research fields share inputs – including funding sources and talent pools – and outputs – such as publications, patents, and trained people. The work that turns one into the other happens under similar "rules of the game." Some rules, like those governing federal grant review or conflict of interest are more formal. Others, like expectations about what constitutes a strong dissertation, or an important finding are more informal. Research fields sometimes share broad logics of action[7] like the sensibility – "a rough sense of direction and an imperative to 'get on with it'" – that some attribute to AI.[8]

They involve many of the same players. It would come as little surprise if cybersecurity and AI were both defined by work done and people trained at the same 30-50 universities, by the employees of a similar number of firms and the investments of a much smaller group of funders. Perhaps more surprisingly, many of the same organizations are likely to be central to synthetic biology or advanced materials research. The organizations that help shape different fields certainly will not be identical, but substantial overlap is common and leads to similarity.

Careers work in similar ways. The most successful people in each field move across these locations over their careers, collaborate with one another, hire each other's students and trainees, build upon, critique, or formally review each other's work. They serve together on the program, prize, and study committees that help set research agendas and define both the formal and informal rules of the game. The best students will often gravitate to the most central institutions and well-connected researchers as will many of the most attractive employers. These dynamics pose significant challenges for efforts to expand equity and broaden participation, but they also drive similarities in how research fields operate that we can leverage for classification.

Conceptualizing technology areas as research fields and emphasizing points of similarity provides a solid, general basis for designing, training, validating, and explaining a "technology agnostic" classificatory model. Identified points of divergence, in contrast, provide leverage to capture salient differences in "technology-specific" model inputs. Anchoring technology specific inputs and a technology agnostic model on a common, fairly intuitive conceptual framework offers a nice balance of flexibility, generalizability and explicability that serves the needs of this use case very nicely.

Shift from documents to people to characterize fields.

Speaking generically, most current research classifications:

- begin with a document corpus,
- extract representations (e.g. topics) of the work they report from some portion of their text,
- cluster those representations in some abstract "knowledge-space,"
- use some algorithmically identified subset of representations to characterize a field.

This conception of a "field" is fundamentally different than our social and organizational definition. It makes social organization the outcome of relationships among representations of ideas extracted from documents. It obscures many sources of commonality and stability that enable generalizable tools and clear measures of change. It increases sensitivity to rapid alterations in technology, substance, or terminology. These problems are amplified at the frontiers of fast-moving multidisciplinary research areas.

We reverse the traditional logic and begin with social organization to identify people working in a field. Rather than saying AI researchers are those who work on some specified set of topics, we determine the topics that constitute AI at a particular time from the portfolio of work produced by AI researchers. Our "people-centric" classification follows the "operational definition" proposed by the AI 100 Year Study Panel: **"AI can also be defined by what AI researchers do."**[8]

Staring with people has four appealing features. It:

- is robust to rapid terminological, methodological, and content shifts;
- allows clear measurements of change in a field's content because people shift much more slowly than the topics they study;
- leverages network theory and measures to operationalize key dimensions of group membership, which supports a clear, explicable approach to data and model design; and
- supports a framework and data architecture for industry classification that directly connects specific research investments to jobs and employers.

Shifting from documents to people is the fundamental move that distinguishes "industries of ideas" classification from what has come before.

Follow the people to identify the industries.

Papers, patents, and other documents are certainly important, but putting people first is very useful. It aids with the practical necessities of data linkage and system building; addresses the key role tacit knowledge plays in work on research frontiers;[9,10] and, by doing both, provides empirical and conceptual routes to connect investments in specific research fields to outcomes in clearly related industries that aren't identifiable by standard means. Put simply, people are the primary output of research investments.

Empirically, we use people's careers to integrate a linked data architecture that reaches from grants through university HR records to state workforce and employer information. Conceptually, we treat employers' decisions to hire people trained through those research investments as concrete, often costly, commitments to the continued development or application of technologies related to the initial research investments.[11] **If AI is what AI researchers do. AI industries are those where employers seek and pay to hire AI researchers.**

As Paul Romer has noted: "Universities produce both papers and people. People with specialized problem-solving skills are the essential input into the discovery process, most of which takes place in the private sector. People with these skills are fuel that fires the innovation engine."

I have made a similar argument about how research universities make distinctive contributions to the public good.[12] The framework outlined here is precisely aligned with the conceptual logic that governed the design of IRIS, the UMETRICS data[13] and ongoing data linkages to administrative workforce data maintained by states and shared with federal agencies. Figure 2 presents an overview of the entire prototype system we are building for two research fields, AI and Electric Vehicles (EV), in a single state, Ohio.

The research field classification that is the primary input to this system is our focus here.

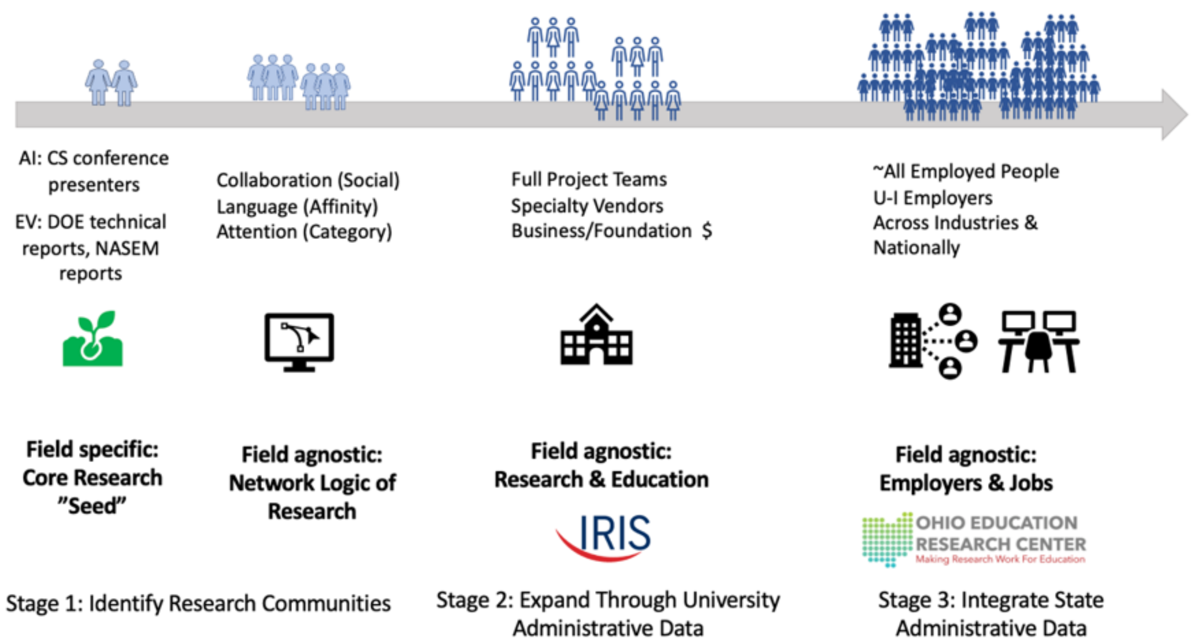

Figure 2. Industry of Ideas Data System

However, it is worth talking briefly about the other data this system will generate. The key challenge of any research classification that draws primarily on bibliometric data is that no matter how good it is it will miss people who have relevant skills developed through research but do not appear as authors on papers. That is quite common either because their roles do not generally lead to authorship (for instance in the case of staff), or because they are excluded from author lists.[14]

We rectify that problem by linking data on authors and their grants to transaction level UMETRICS data, which lets us see everyone those grants pay whether they are faculty, staff, undergraduate students, doctoral students, or post-docs. This move dramatically expands our definition of the impact of research investments and also captures a much more extensive "tracer" set of individuals whose later career mobility can allow us to describe job, employer, and workforce implications for newly identified sets of industries.

We classify the industries relevant to research fields using existing data infrastructures maintained by states that contain detailed wage records for all employees who are eligible for unemployment insurance (U-I). Universities are among those employers, so the people - both those identified as authors and those paid on grants who are not - represented in university administrative data can easily be linked to state U-I wage records. Their post-university employment describes the employers that have bid into a relevant research field by hiring research trained people from that field. The people they hire can be connected back to concrete research investments associated with specific technologies.

Once the employers associated with a research field are characterized, all the jobs they support and all the other employers and jobs in their traditionally categorized industries can be richly described. People and their careers provide the conceptual and empirical throughline that makes both system design and a measurement approach that can address CSA mandates possible. With appropriate institutional arrangements and partnerships, linkages to other types of state data such as K-12, higher education or social service information may become possible.

A strong classifier is both an essential part of the larger data system we need to address the measurement challenges we now face and a tool that can be used with data from that system to expand our analytic horizons. The results could simultaneously push knowledge in several fields forward *and* generate high value products for many stakeholders, creating a virtuous cycle and accelerating growth to scale.

Implementing and Improving a Classification Strategy.

What should a field classification system be able to do?

A good field classification strategy should apply to multiple technology areas and be broadly legible to non-specialists. More specifically a strong classification approach should do four things:

1. <u>Identify core technical contributors.</u> In AI, it should confidently classify researchers developing models and techniques at the frontiers of the field's current technical core.
2. <u>Identify researchers applying tools from the field in diverse domains</u>. AI tools and models are used for many purposes across biomedicine, genetics, physical science, materials, information and social science, engineering, and other areas. Fully measuring the impact of research investments and effective industry classification both require reliable capture of sophisticated applications in many substantive areas. For some CSA technology areas, including AI, it may also be useful to identify researchers whose work addresses relevant ethical, legal, and social (ELS) issues.
3. <u>Classify individuals.</u> One core task of a classification system is, well, classification. However, the first two needs highlighted above suggest the benefits of a categorical (e.g. core, domain, ELS) rather than a binary approach.
4. <u>Provide quantitative estimates.</u> Given the cross-cutting nature of many CSA technology areas, a regression approach that estimates the intensity of a given researcher's engagement with AI or a continuous measure of proximity to the field's technical core may prove more useful than a classification. For instance, researchers exclusively focused on technical improvements to deep learning models should be very closely associated with AI. Researchers who improve such models and work on non-AI substantive problems or those who address substantive questions with AI and non-AI methods might be less strongly associated with the field. Continuous estimates could be

used to construct measures for inferential analysis (e.g., weighted averages) or to allow analysts to vary thresholds for different use cases. If a reasonable measure of “direction” could be incorporated into a proximity estimate (e.g. moderately distant from the AI core in the direction of social science) to create a vector, even more nuanced analyses might be possible.

Step 1: Defining a “seed set” of researchers.

The general approach has two steps. First, consultations with domain experts help identify a plausible “seed set” of researchers. That seed group should offer an easily definable, face valid input or “starting point” for a more general, model-based second step, but it need not be comprehensive. It should take the distinctive social and organizational characteristics of the research field into account.

This technology specific seed definition will typically emphasize the core technical areas of the field, but it should also include relevant domain researchers who are highly proximate to that core. This is not intended to be an exhaustive and exclusive “gold standard.” Rather, we seek a plausible initial definition of the field’s core participants. Including people who might be “network bridges” into potentially far-flung domains where technologies developed in or near the core are commonly applied ensures that modeling can identify researchers applying AI in disparate substantive areas.

Table 1. AI Conferences That Define Initial Researcher "Seed Set"

Conference Title
AAAI Conference on Artificial Intelligence
ACM Conference on Recommender Systems
Association for Computational Linguistics
Computer Vision and Pattern Recognition
Conference on Lexical and Computational Semantics
Conference on Uncertainty in Artificial Intelligence
Empirical Methods in Natural Language Processing
IEEE International Conference on Data Mining
International ACM SIGIR Conference on Research and Development in Information Retrieval
International Conference on Computer Vision
International Conference on Information and Knowledge Management
International Conference on Learning Representations
International Conference on Machine Learning
International Conference on Principles of Knowledge Representation and Reasoning
International Joint Conference on Artificial Intelligence
International Joint Conference on Natural Language Processing
Joint International Conference on Computational Linguistics
Neural Information Processing Systems
Robotics: Science and Systems
Special Inerest Group on Knowledge Discovery and Data Mining (SIGKDD)
Winter Conference on Applications of Computer Vision

For AI, we defined an initial seed set based on presentations at prominent AI and Machine Learning (ML) conferences. Table 1 presents a list of those conferences. In collaboration with analysts at Elsevier, we identified people who had authored on at least one paper presented at any of these meetings since 2010. That highly international group of researchers included 94,235 individuals.

Step #2: Model-based expansion

Seed set researchers and associated data about their publication and grant histories, co-authors, and affiliations provide the input for a technology agnostic model in the second stage of our classification approach. AI seed researchers were used as a labelled subset of all authors in the Elsevier publication corpus. We implemented a shallow Bayesian label propagation model[15] on the unweighted co-authorship graph to classify unlabeled authors as potential AI researchers. The model, which was implemented by collaborators at Johns Hopkins University, identified an additional 154,096 unique AI authors. It also successfully labelled 21,803 authors who were in the seed set but appeared in full co-authorship networks as unlabeled nodes prior to their first relevant conference presentation. In other words, more than 23% of seed set authors were publishing actively before their first AI conference paper, were captured by the label propagation model, and later observed as an author on an AI conference paper.

Seed set and model identified authors included 248,331 unique researchers. About 37% (97,379) listed U.S. affiliations on any of the 1.96 million research publications they collectively authored. This represents an initial cut at a person-centric, social and organizational approach to AI classification that can potentially meet the first three requirements I lay out for a classification strategy.

These nearly 250,000 people offer a reasonable first approximation of the global set of researchers likely to possess AI-relevant skills. Such skills could make them, their students, post-docs, technical staff and close collaborators, attractive candidates for jobs tasked with developing or applying AI tools and technologies to existing or new products, services, or business processes. Due to our definition of the seed set and our model design, this group will include both core technical AI scientists and domain application researchers from many fields.

Using them as a starting point for an industry classification that depends on hiring as a marker of employer engagement with AI will thus identify firms active in a wide range of traditionally defined industries. By doing so we address one of the core challenges that animates this conference, the identification of "industries" that aren't visible in traditional classifications. Nevertheless, this group faces a key limitation in that it only includes people who authored papers published in venues indexed by Elsevier. This is why linking the results of a field classifier to UMETRICS data before proceeding to industry classification and associated measurement adds important substantive and technical components to the larger measurement system.

Validation and model improvement work is ongoing along with a test of the entire process on a second, very different technology, Electric Vehicles (EV). Early results suggest two important needs for refinement. First, this initial model may have difficulty identifying technical AI researchers from outside computer science and computer scientists developing ML as a method to address other substantive issues who do not attend AI conferences. Second, and more importantly, the model is less effective at identifying domain application researchers. Accuracy seems to fall off most steeply in biomedical areas.

Both challenges may result in part from model inputs. Our definition of the seed set may need to be expanded with a particular emphasis on including more researchers developing and applying AI methods to substantive problems, especially in biomedicine. Here I address the possibility that the challenges result from our initial label propagation model's relative algorithmic and empirical simplicity. I suspect that a shallow model relying solely on the co-authorship graph imposes unnecessary limitations. As an alternative, I sketch a more thorough network conceptualization of group membership and outline a more sophisticated semi-supervised deep learning model.

Conceptual Approach to Improving the Classification Model.
Beginning with people situated in a socially defined research field allows us to treat classification as, fundamentally, a problem of establishing individual membership in a social group. The basic question we want to answer is "given some set of information about a person, X, can we say X belongs (or is more or less closely related) to a group, Y, for which we have comparable information about a set of known members." Here X is an author on a scientific paper whose membership status relative to Y is unknown and Y is AI. Our seed set authors are the labelled AI representatives. The comparable information is a feature set drawn from bibliometric data.

Questions about groups and their members have been fundamental to Sociology for more than a century.[16] More recently, network science has dramatically expanded relevant methods and measures.[17–20] This classification problem is an effort to establish group membership using networks relevant to socially defined research fields. Viewing it that way offers well-established routes to justify and explain data and model design choices that align closely with our overall conceptual approach. Such theoretical and methodological integration contributes to the clarity, generality and legibility of the field classification and the larger measurement system of which it is an essential part.

In the very broadest terms, group membership is a function of three general mechanisms.

- Social proximity – Traditionally, this is a literal question about physical proximity or kinship. But proximity can be framed socially in terms of non-familial relationships regardless of physical co-presence or consanguinity. When they are applied to graphs of social relationships among people, this sense of social proximity animates most current community finding algorithms.[21–24]
- Affinity – Group membership can be based on shared likes and dislikes. People who like, talk about, and attend to similar things in similar ways will often identify as members of a group and act from those identities.[25,26]
- Signaling – Sometimes group membership is as much about others' perceptions as it is about an individual's actions or beliefs. People's actions can send signals to others (intentionally or not) about their membership in groups. Others can ascribe membership status (erroneously or not) to individuals based on things they infer from what they observe.[27–29]

Typically, group membership means some combination of: X is socially close to other members of Y; X feels and demonstrates a connection to members Y based on shared interests, language, and activities; and X makes (often symbolic) claims to membership in Y that others recognize and can accept or reject. This yields a few simple propositions. A researcher is an AI scientist if:

- They are professionally connected to known AI scientists (social proximity)
- They work on topics and use language similar to known AI scientists (affinity)
- Their professional activities allow others to infer they are AI scientists (signaling)

Operationalizing Dimensions of Group Membership

Our current model focuses exclusively on a single membership dimension, social proximity, measured via co-authorship. This unitary focus might limit the model's ability to classify domain application researchers whose social distance from core technical AI could limit the likelihood they will co-author with seed set scientists. Social proximity is arguably the most direct measure of membership but may be too limiting for this use case.

Expanding model input data with features designed to operationalize affinity and signaling could address that. Consider two concrete examples.

Measurement Example: Affinity

Affinity, framed in terms of language and interests, could be measured with vector embeddings from a pre-trained language model. SciBert,[30] a transformer model trained on a large scientific corpus, could be fine-tuned, and used to position labelled and unlabeled authors in a "conceptual space" based on the text of their published abstracts. Distances among them could be calculated in many ways either in cross-section or longitudinally.[31] Smaller distances would indicate greater affinity.

Pairwise distances among researchers could be represented as a valued network and treated as another route for label propagation. Affinity connections are likely to bridge gaps or collapse distances in co-authorship networks, as many people who are unlikely to ever work together directly study very similar things in very similar ways. So, adding a second network dimension focused on affinity has significant potential to improve model performance.

Moreover, recent has work demonstrates the value of fine-tuning multiple versions of transformer models with text from different time periods to capture changes in the represented spaces themselves.[32] This measurement approach could capture shifts in individual activities and larger alterations to the field as a whole. Peoples' interests might change over time, but the landscape on which they pursue their interests is also dynamic.[33]

Measurement Example 2: Signaling

Signaling can also be operationalized in numerous ways. One example that aligns nicely with our social definition of fields, relies on publication venues as markers of participation in particular research areas. The choices scientists make about where to publish their work can be orthogonal to the substantive and methodological content of that work. They are also decisions about what intellectual communities to take part in and what kinds of professional identities to build and maintain. If, as I often do, I have a paper that could fit in a policy journal, a sociology journal, or a management journal, deciding where to send it is about the field I want to engage with and the audience I want to reach. It is a decision about the signal I want to send about how the paper's contents should be understood. Researchers often read CVs with this kind of signaling in mind because where one chooses to submit and succeeds in placing papers tells insiders much about the kind of researcher one is or seeks to be.

Since scientific publication is a two-sided affair where editors and reviewers select manuscripts in part based on "fit," publication venue offers a clear example of a signaling mechanism. As a result, we might define group membership in terms of the degree to which a pair of scientists' scholarly works appear in overlapping venues. Greater similarity, by virtue of choices about submission and acceptance, equates to stronger signals of membership in the same group. Any one of several specific measures calculated on a weighted or unweighted author by journal matrix in cross-section or longitudinally could be used. All of them could yield another valued network linking researchers.

A Multidimensional Network (with light formalization)

For illustrative purposes consider a simple cross-sectional data structure that could be implemented for any research field. The overall data structure would be a multidimensional graph eminently suited to tensor representation.[34–37]

Let the multidimensional graph $\mathcal{G} = \{\mathcal{V}, \mathcal{D}\}$ where $\mathcal{V}$ is a set of nodes $\{v_i \ldots v_N\}$ and multiple sets of edges $\{\mathcal{E}_1 . \; . \; . \mathcal{E}_D\}$ connect those nodes in $\mathcal{D}$ dimensions. Edges in each dimension can be represented by a set of adjacency matrices $\mathbf{A_1} \; . \; . \; . \; \mathbf{A_D}$ where $\mathbf{A_d}[i, j]$ represents the connection in dimension *d* between node *i* and node *j*, which can be binary or valued. Nodes can also be represented by a feature vector, **X**, which, for the sake of simplicity I limit for now to a single binary label which indicates whether a given scientist was identified as a member of the "seed set" of AI researchers.

Assume that nodes represent individual researchers and edges represent connections defined on three dimensions: (1) Social proximity, $\mathbf{A_1}[i,j] = 1$ when scientist *i* has co-authored with scientist *j*, otherwise, 0; (2) Affinity, $\mathbf{A_2}[i,j]$ = some real number value representing the conceptual similarity between scientist *i* and scientist *j* defined via a vector embedding; and (3) Signaling, $\mathbf{A_3}[i,j]$ = some real number value representing the degree of commonality in publication venues between scientist *i* and scientist *j.* Figure 3 presents a schematic and hypothetical representation of this kind of graph.

Without going into too much detail, consider scientist 1, who is isolated in $\mathbf{A_1}$ by virtue of having no co-authoring relationships. In our initial model, which relied solely on social proximity, there is no path by which the AI label associated with node 3, an identified AI scientist, can propagate. In this multidimensional data structure, however, there are multiple direct and indirect paths for propagation through the signaling ($\mathbf{A_3}[1,3] > 0$) and affinity dimensions.

Likewise, nuance is added even when co-authorship ties are present. Consider the multiplex relationship connecting scientist 3 and scientist 5 ($\mathbf{A_1}[3,5] = 1$, $\mathbf{A_2}[3,5] > 0$). The pattern of relationships these scientists indicates that they co-author together and that their abstracts use similar language but that (absent their co-authored work, which we would likely exclude in defining signaling ties) they publish in non-overlapping journals. It seems highly likely that a sophisticated model would classify scientist 5 as an AI researcher. However,

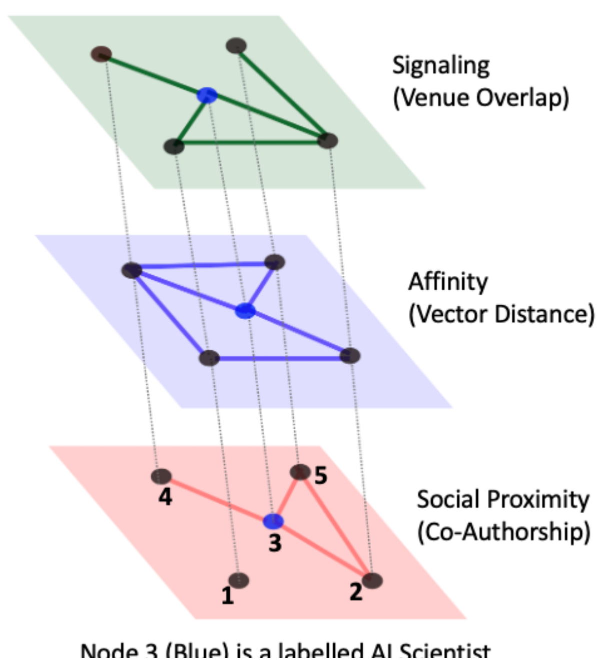

Figure 3. Multidimensional Network for Field Classification

this qualitative pattern of ties might intuitively suggest that scientist 5 is a domain researcher whose application presents sophisticated enough challenges to serve as a case for AI tool development rather than a core AI researcher. Similarly, scientists 4 and 2, who have co-authored with scientist 3 and who share both affinity and signaling relationships with them, seem highly likely to be pursuing research closer to the core of the field even though they are not observed in the seed set.

Modelling Considerations.

This data structure is tailor made for a deep learning model that operates on graphs. Graph Neural Networks (GNNs) offer many possibilities for model designs that can take multi-step network neighborhoods, message passing, attention mechanisms, node and edge features into account using a growing panoply of operators, including some tailored specifically for multidimensional data.[37–43]

While a first-pass model should focus on simple classification, the base architecture could readily yield a quantitative estimate of individual scientists' proximity to the fields' core via a regression task. This kind of model applied to these kinds of data seems well equipped to satisfy all four desiderata I suggest for a strong field classification. More importantly, a unified social scientific conceptual framework offers a plausible basis for expecting that it could be generalized across CSA technology areas.

Deep learning models are basically opaque to explanation.[44–46] Nevertheless, the intuitions suggested by Figure 3 provide useful "hooks" for helping lay audiences make sense of our logic and ideally the results. The bedrock empirical insights we derive from well-established theory and findings can help concretize the workings of complex models. In what follows I draw heavily on (and entirely borrow formalizations from) a recent review by Bronstein and colleagues,[40] who partition spatial GNNs into three basic "flavors." I consider two, the convolutional and attentional, in some detail in order: (1) to lay the foundation for thinking about model design in this case; and (2) to suggest a general strategy for model design that helps make this kind of work a bit more legible for non-expert audiences.

All spatial GNNs use local network structure to update the "state" of a focal node, *u*, by appeal the characteristics of the other nodes, *v,* to which it connects. The key idea is the network neighborhood, $\mathcal{N}_u = \{v \mid (u,\ v) \in\ \mathcal{E}\}$. A given GNN layer, essentially, iterates across the neighborhoods of every node in a network, updating their "states" based on the characteristics of their neighbors and then passing the new state information to a subsequent model layer for further processing.

For a network with both a node feature vector **X** and an adjacency matrix **A** the GNN layer "constructs a *permutation equivariant* function **F(X, A)** by applying shared *permutation invariant* functions $\phi\left(x_u, X_{\mathcal{N}_u}\right)$, over local neighborhoods. (Ref. 40: p 78)" By doing so, the GNN layer preserves some of the structural features of the larger network embedded in the adjacency matrix. It uses those features to weight the effects nodes have upon each other, making use of the complex interdependencies that are the hallmark of network data. ϕ is often called the "updating" function.

I dig into common approaches to ϕ to consider how a GNN classifier anchored on our conceptual and empirical framework might be rendered legible to the audiences that need to understand and use its results. What ϕ does, at the micro level of a single node and its immediate neighbors is somewhat akin to label propagation, which cashes out to providing an answer to our general question about membership in a group. In terms of this section's focus, that question could be restated as: Given some mix of connections to nodes known to be AI researchers, how might differently configured GNN layers update information about whether a given scientist is also an AI research? What we are interested in, then, is h_u, the "state" of the node u. In our case, that state might be something called "AI-ness." Which brings us back to the some of the "flavors" of spatial GNNs defined by their general approach to ϕ.

Relying on graph structure: Convolutional approaches

Convolutional GNN layers would address the question of a given scientists' "AI-ness", h_u in the following generic fashion.

(1) $h_u = \phi\left(x_{u,} \bigoplus_{v \in \mathcal{N}_u} c_{vu} \psi(x_v)\right)$

Ignore ϕ and ψ for now. They are "model appropriate" transformations, generally affine, that are learned through the training process and often modified by activation functions such as the rectified linear unit (ReLU). What is important to understand to get an empirical taste of the approach is the aggregation function ($\bigoplus_{v \in \mathcal{N}_u}$) and c_{vu}. The aggregation function is what determines how you get from many neighbors each of which has some state (in our simple case, binary) of AI-ness to an updated state for u. $\bigoplus$ must be permutation invariant, which means its value will not change if you switch around the order of the things operates on. Sums, averages, and maximums are all common aggregation functions. We are going to treat $\bigoplus$ as a sum. The other term, c_{vu}, is what determines how important each neighbor is for updating the state of u. In convolutional approaches that term is a constant, which is generally defined by the the adjacency matrix **A,** sometimes after transformation**.**

Stripped of all the gnarly but very useful mechanics of machine learning, what this resolves to in our simple case is a really basic empirical intuition. A scientists' AI-ness is a function of the sum of the AI-ness of the other scientists to whom they are connected. Where our seed set labels are binary, this basically says a scientist has greater AI-ness when they are connected to more researchers who are themselves AI scientists. The more AI researchers you co-author with, share affinities with, and publish in similar venues with the more likely you are to be an AI scientist yourself.

There are complications. We have conceptual reasons to believe that the different dimensions have different implications for establishing AI-ness. None of that matters to any deep learning model we might eventually train. Right now, we cannot really know how such a model will learn whatever it learns. Smarter people than me are working the problem. But there are theory-based changes we could make to address some of our intuitions.

For instance, we could modify c_{vu} to deal with features of edges such as their values. If we binarized all three dimensions, we might simply weight the effect of a neighbor's AI-ness on u by the number of dimensions along which the pair were connected. Recall Figure 3 and the difference between scientist 1 (who had one signaling connection to scientist 3, the known AI researcher) and scientists 2 and 4 who each connected to 3 in all three dimensions. More complicated approaches could allow continuous values for all ties and apply further transformations to those.

Regardless, convolutional approaches work primarily from network structure to update node states. In our case that means who you are connected to and whether they are part of the seed set is what matters.

Dynamic weights based on attention.

A somewhat more complicated flavor of GNN layer is *attentional*.[39]

(2) $h_u = \phi\left(x_{u,} \bigoplus_{v \in \mathcal{N}_u} a(x_u, x_v)\psi(x_v)\right)$

The only change here is the replacement of c_{vu} with $a(x_u,x_v)$, a self-attention mechanism.[*] Instead of treating *v's* influence on *u* as a constant, driven more or less solely by observed ties, this approach introduces a learnable parameter that calculates "importance coefficients," $a_{uv} = a(x_u, x_v)$, which weight that influence by the features of neighbors. The underlying intuition might be summarized, in the social proximity dimension of our network, by the phrase "not all co-authors are created equal." Some partners have characteristics that make them more important or salient to *u's* state than others. The attention parameter creates a variable means to accommodate such differences in an aggregation function.

Reducing this general description to a more specific empirical intuition in the highly simplified (single, binary node feature) case described above is trivial. So, imagine two possible complications. The first adds additional features to the node vector, **X**. The second builds on the logic of the signaling dimension, $\mathbf{A_3}$, itself. In a research field, such as AI, the core empirical intuition is that *u* might be connected to an AI scientist via one or more of our dimensions but either be unaware of or uninterested in the fact of that connection.

Absent a reason for *u* to pay attention to an AI co-author, the fact of co-authorship alone may have little effect on their state. This might be especially true in the fields where team sizes tend to be large, or where people tend to publish many papers with a wide array of co-authors. Consider high energy physics or population genomics. Both are fields where AI tools are increasingly broadly used. Both are also fields where papers can routinely have hundreds to thousands of co-authors. The fact of co-authorship with an AI scientist on such a paper may have very little actual bearing on the AI-ness of any given physicist or genetics researcher. For what it is worth, things like norms about team size are exactly the kinds of research field characteristics that could be captured by the social and organizational approach we propose and that might need to be accommodated in seed set definitions.

In the network that interests us, a node's state change is driven by the intertwined mechanisms of social influence (people become more like those to whom they are connected) and homophily (people are more likely to connect to partners who they are already similar to). The idea that being connected to an AI scientist indicates one is more likely to be an AI scientist oneself either presumes that: (a) when non-AI researchers collaborate with AI researchers their

[*] In the special case of a complete graph, which, if no threshold is applied, will be the case with the affinity dimension, $\mathbf{A_2}$, as we have defined it. This generic equation reduces to the forward pass of a transformer, which is typically driven by a multiheaded attention mechanism.[47,40]

AI-ness increases as a result of the interaction (social influence); or (b) that AI scientists are more likely to collaborate with one another and thus observing a tie reveals an otherwise difficult to observe categorical similarity (homophily).

These two mechanisms are extremely difficult to tease apart,[48,49] but both depend on social forms of attention. Whether an existing similarity draws one to connect to collaborator or an existing connection increases one's similarity to them, people must: (1) be *aware* of the characteristics they share; and (2) those characteristics must be *salient* enough to factor in decision-making.[50,51] Two questions follow: (1) what might lead researchers to be more aware of the AI-ness of their network neighbors? (2) what kinds of things are likely to make some network neighbors more salient than others?

The answers may vary across our three network dimensions. The bar for awareness and salience seems likely to be higher for the affinity dimension, where the sheer fact that two scientists write in fashions that position them near each other in a complex vector space offers no assurance that they will know of each other or have ever read each other's work. Regardless, either additional node features or a refined conception of signaling can offer examples of potential answers.

Consider just one class of node features that might have traction. Higher profile neighbors are both noticeable and salient. Though it would offer no purchase on actual GNN outputs, an illustrative attention mechanism based on scientific visibility suggests that more highly cited neighbors, neighbors who have won high status awards, or neighbors affiliated with high visibility institutions or programs would exert greater influence *u*'s state than others. As I started work on this paper, I used all these markers to help guide my attention as I immersed myself in a large, complex, and wholly unfamiliar literature.

Once again, the benefits of a broadly social conception of research fields are apparent. For instance, we do not need to determine exactly which institutions have the highest status in a field to recognize that institutional affiliations are likely to be important attention features. We don't even need to attend to how the ranks of institutions change across fields. We simply need to know that all fields have status hierarchies and that they shape researcher attention. Knowing that, we construct input data that will increase our confidence and the legibility of a model design that includes attention mechanisms.

Alternatively, we might return to the signaling dimension of our network. The signaling mechanism, fundamentally, relies on attention. Recall the intuition behind our proposed measure. Researchers try to place work in venues that reach intellectual communities to which they belong or hope to belong. Publishing in those venues also, typically, leads to reviewing for them, which focuses limited individual attention even more tightly on a small set of journals and increases awareness of others publishing or trying to publish in them. As careers progress, ad hoc reviewing turns into program committee or editorial board memberships and sometimes program chair or editorial roles. All these transitions further focus attention and deepen both the visibility and the salience of others who publish in those places.

These generic social dynamics are common in contemporary research fields. They suggest that $\mathbf{A_3}$, the signaling adjacency matrix, measures signaling precisely because it represents a general structure of attention. This, for what it is worth, is entirely consistent with social science research that examines networks, categories and signals in a wide range of fields.[52–57] More concretely, this requires we attend to edge features in the signaling dimension as we think through the empirical logic of attention layers.[41,42]

There are a few takeaways. First, paying attention to attention mechanisms requires us to think through: (a) the conceptual relationship between the social mechanisms (homophily, influence) that underpin our intuitions about state updating at the level of nodes and their neighborhoods; (b) the signaling dimension of our multi-dimensional graph and its implications for attention in the social organization of research fields; and (c) the way both should shape choices about what node and edge features are necessary to maintain some degree of conceptual, if not actual, explicability. Second, the discussion so far, which has focused solely on trying to make layer by layer sense of baseline empirical intuitions for different GNN methods already suggests implications for model design that, by and large, follow from the general conception of research fields we propose.

Musing about model design

One of the many tricks of designing and training a working model appears to lie in stringing together an appropriate set of layers tuned to accomplish a particular task. Such a design is beyond the scope of this white paper, and, frankly, beyond my skills. But that will not stop me from speculating and asking that anyone who might be interested in collaborating or simply in correcting me reach out.

With necessary disclaimers out of the way, the discussion above suggests a model design based on a “block” composed of three layers. Our multi-dimensional network inputs would include three valued adjacency matrices ($\mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3}$) corresponding to the network dimensions (proximity, affinity, and signaling), as well as a vector of node features (**X**) that comprised of scientist level measures defined by appeal to the social and organizational field definition and the task we want the model to perform. We want an effective, technology agnostic model and training data that accord with a conceptual framework that can be made legible to non-specialists. These data would provide the initial input states for a convolutional layer, which would update those states based on the structure of connection in each node’s 1-hop neighborhood and pass the resulting, updated states on to an attentional layer emphasizing node features, which in turn would pass updated states to a second attentional layer focused on edge features.

This is where, as I understand it, the magic of deep learning models begins to kick in. The learnable transformations we have been resolutely ignoring (ϕ, ψ) are updated across layers through the training process as are the weights that result from each layer. As the input data pass through each layer of the model, the output of a prior layer serves as the input to a subsequent layer and through training the sequential application of different “flavors” of GNNs results in more and more refined predictions.

The three-layer block we suggest could thus be understood to yield a progressively more fine-grained aggregation of neighboring node states starting with the coarsest measure (connection) and proceeding through two levels of attentional weighting. Depending on specific decisions about input data (particularly the values ascribed to edges and the features included in node vectors) different features of the broad social conception of group membership that underpins this proposed architecture might be emphasized. Equally importantly, the application of multiple blocks could allow learned weights to reflect broader neighborhoods, so a second “block” might be understood to encompass weights aggregated at network degree 2 (collaborators of collaborators), a third at network degree 3 (collaborators of collaborators of collaborators) etc. Of course, many more details need to be worked out, but a general model architecture like that sketched in Figure 4 at least offers a starting point for thinking.

Whether the final details of the “technology agnostic” model end up following this set of suggestions is not particularly important. What is important is that any model architecture included in the larger industries of ideas data system align with the general principles that helped us articulate this one.

- Emerging technology areas are research fields defined in social and organizational terms
- People define the boundaries and contents of research fields
- Field Classification is a two-stage process where
    - Domain experts help define a “technology-specific” researcher seed set that includes both core technical contributors and people pursuing domain applications in a range of areas
    - Social theory and network measures inform “technology-agnostic” data structures and a semi-supervised GNN model to expand the seed set via classification and regression tasks
- The results of that field classification are essential inputs to an integrated “people-centric” data system that:
    - uses UMETRICS data to expand from authors to all grant employed research personnel working on a particular technology;
    - links both authors and non-author university grant employees to state administrative (U-I) wage data;
    - uses employee mobility identify specific employers who have hired people trained on grants relevant to that technology;
    - classifies industries by treating hiring as evidence that employers have “bid in” to pursue work relevant to the application or development of the technology in their products, services, or processes;
    - uses “hiring bids” as indications that employers’ traditionally defined industries, the other employers in them, the jobs and the people who hold them should be included in efforts to describe and estimate the impacts of research investments;

Following these general principles to classify fields and industries across CSA technology areas supports a measurement system with all the components needed to rigorously address the challenges posed by the CHIPS and Science Act.

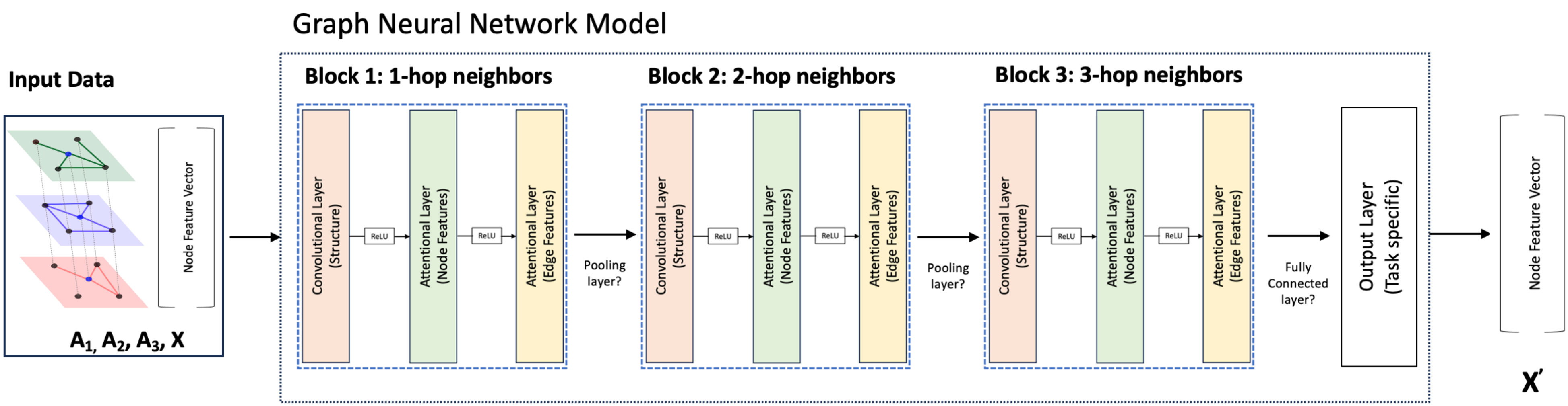

Figure 4. Sketch of a GNN Architecture

Brief thoughts on AI-relevant extensions.

Integrating a social and organizational classification strategy with the larger "industries of ideas" framework and data system also creates immediate opportunities to address additional AI-specific measurement concerns.

Measuring the impact of "platform" investments.
One of the major concerns of the National Artificial Intelligence Research Resource (NAIRR) task force report is the availability and accessibility of compute and testbed resources. But the impact of such platform investments can be difficult to evaluate. The users of shared platforms typically hail from many fields and institutions. The platforms themselves generally collect information only on those who use them. This gap makes identifying a comparison group impossible, creating a roadblock for assessment. The prototype data system described here, in partnership with compute and testbed operators, is uniquely suited to address this challenge.

UMETRICS data on the complete set of grant funded people working at universities coupled with employment information derived from state U-I wage data offers many possibilities for developing counterfactuals. Indicators and quantitative measures of "AI-ness" drawn from a strong field classification add more. What is needed once such a system is constructed is simply the ability to link user information for compute and testbed resources to the larger data infrastructure at the individual level. That integration enables many research designs to assess scientific, technological, and workforce impact and could potentially guide more effective development and use of shared AI resources.

Software and model development as an impact of research investment

One of the impressive analyses in the AI Index tracks technical improvements in model development. This is an essential component of the field that I simply don't address. It could be systematically examined by integrating another type of data into the industries of ideas ecosystem.

The Institute for Research on Innovation & Science (IRIS), which I co-founded with Julia Lane and Bruce Weinberg, is currently working on a pilot project to describe the relationship between research funding and the development and use of research software. We extract software mentions[58] from the text of scientific papers then link named packages, the papers and authors that use them to software repository metadata and UMETRICS. These linked data connect research investments, research publications and software tools. As is the case with the industries of ideas framework, those linkages "follow the people." By virtue of connection to repository metadata, they also provide many new data points about the code itself, its use and process of development, whether and how it is maintained and updated and other features.

Research software that implements new AI models, like other intermediate research products such as datasets,[59] is an important research output that not often systematically studied. In AI -- where most papers are openly accessible in full text formats, where two major Python packages (PyTorch and Tensorflow) are the primary basis for development, expansion, and application of new machine learning models, where standard benchmark datasets are commonly used and referenced, and where code is generally shared upon release of working or conference papers -- linking code repository metadata into a unified data architecture could offer immense value.

First, such data could expand work like that Stanford's Human-Centered Artificial Intelligence institute already does to track the characteristics and use of new models. Second, it might allow us to add entire new classes of inputs, outputs, and, potentially, people to the field and industry classifications we describe here. Third, we might expand affinity and signaling dimensions for field classification input data by appeal to use of the same tools or datasets. Finally, to the extent that code developers and maintainers do not completely overlap with authors on papers, we might be able to identify yet another group of people touched by AI researcher investments whose careers could aid in industry classification. While such expansions might not be generalizable across all CSA fields, they seem likely to be particularly valuable for those where code and data are both core outcomes and enabling tools such as AI, cybersecurity, and distributed ledger technologies.

Modelling the dynamics of regional innovation ecosystems

The CHIPS and Science Act is one of several substantial "place-based" investments made under the Biden administration. Some, including those in the CSA, are larger than similar Great Society and even New Deal programs. Associated measurement challenges reach beyond job and employer effects to holistic program impact on regional innovation ecosystem dynamics and outcomes.

Two features that distinguish successful and unsuccessful regional innovation ecosystems are: (1) the founding of "second generation" start-up companies;[60–64] and (2) the formation of dense cross-employer mobility networks. Early research on the success of Silicon Valley highlighted both, citing the "Fairchildren" firms that grew out of Fairchild Semiconductor and the ability to change jobs without changing carpools in explanations of to the Valley's resilience relative Boston's Route 128.[65,66] Learning by hiring remains a key source of competitive advantage in technology intensive regional ecosystems.[67,68]

A high-profile example from contemporary Bay Area AI clearly illustrates the point. Consider "Attention is all you need,"[47] a 2017 paper that helped spark the explosion of large language models that has been cited about 105,000 times. It was written by 8 authors at Google. By 2021, all 8 had departed. Searching for them today reveals that they have collectively worked for or founded 9 AI companies since leaving Google. Six of the original 8 authors remain in the Bay Area. Seven of the 9 companies, most of which were founded after the paper's publication, are also located in region. The mobility networks among employers created by such moves and the founding of such "later generation" firms are a large part of what make regions like the Bay Area such successful and resilient technology ecosystems.

But the data and tools needed to systematically (1) assess when and how investments of different sorts might help regions become self-sustaining, (2) support second generation entrepreneurship, or (3) predict how such networks develop have never existed at scale. The data systems sketched here, which match rich, though restricted, workforce date with detailed bibliometric and university information could change that, dramatically. Long time series data for established and emerging regions with and without investments from programs like the Regional Innovation Engines or EDA Technology Hubs allow pre and post-investment analyses.

Graph based deep learning models trained on those types of data could offer a wide range of new insights at the regional level.

Models that treat regions as networks and use graph classification to identify structural features that are leading indicators of regional success offer one interesting possibility. Link prediction tools that focus on the evolution of mobility ties within the region (or into it) offer another. In the latter case, we might envision exciting possibilities for both understanding the regional development around anchor tenants such as universities[12,69,70] and the role that key "on ramp" institutions such as community colleges play in ensuring broad access to jobs created by programs like the Regional Innovation Engines. Combining deep learning models with granular administrative data might go a long way toward cracking questions about educational and workforce capacity and their relationship to regional dynamism. The ability to address such questions would be dramatically increased by inclusion of state maintained higher education or K-12 data in the industries of ideas ecosystem.

If education data could be included in the mix, it might become possible, for the first time, to systematically assess the essential role that skilled technical workers play in creating and sustaining regional success. This kind of analysis would be particularly important for questions about manufacturing capacity and manufacturing-oriented CSA technology areas. In the context of reciprocal partnerships and effective governance the need for strong privacy protections on restricted data might make famously complex and inexplicable models an important selling point.

Works Cited

1. Lane, J. *Democratizing Our Data*. *Democratizing Our Data* (MIT Press, 2020). doi:10.7551/mitpress/11990.001.0001.
2. Chang, W.-Y., Garner, M., Basner, J., Weinberg, B. & Owen-Smith, J. A Linked Data Mosaic for Policy-Relevant Research on Science and Innovation: Value, Transparency, Rigor, and Community. *Harvard Data Science Review* (2022) doi:10.1162/99608f92.1e23fb3f.
3. Kreuter, F., Ghani, R. & Lane, J. Change through Data: A Public Extension Program for Government Employees. *Harvard Data Science Review* **1**, (2019).
4. Lane, J. *Reimagining Labor Market Information: A National Collaborative for Local Workforce Information*. https://www.aei.org/research-products/report/reimagining-labor-market-information-a-national-collaborative-for-local-workforce-information/ . (2023).
5. Powell, W. W., Koput, K. W., White, D. R. & Owen-Smith, J. Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology* **110** (2005).
6. DiMaggio, P. & Powell, W. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* (1983).
7. Friedland R. & Alford, R. Bringing Society Back In: Symbols, Practices and Institutional Contradictions. In *The New Institutionalism in Organizational Analysis* (1991).
8. Artificial Intelligence and life in 2030: The one hundred year study on artificial intelligence (2016)
9. Collins, H. M. Tacit knowledge, trust and the Q of Sapphire. *Social Studies of Science* **31**, (2001).
10. Collins, H. *Tacit and Explicit Knowledge*. (University of Chicago Press, 2013). doi:10.7208/chicago/9780226113821.001.0001.
11. Lane, J. I. *The Industry of Ideas: Measuring How Artificial Intelligence Changes Labor Markets*. https://www.aei.org/research-products/report/the-industry-of-ideas-measuring-how-artificial-intelligence-changes-labor-markets/ (2023).
12. Owen-Smith, J. *Research Universities and the Public Good: Discovery for an Uncertain Future*. (Stanford University Press, 2018).
13. Nicholls, N., Ku, R., Brown, C. & Owen-Smith, J. *Summary Documentation for the IRIS UMETRICS 2022 Data Release.* doi:10.21987/df2a-ha30.
14. Ross, M. B. *et al.* Women are credited less in science than men. *Nature* **608**, 135–145 (2022).
15. Yamaguchi, Y. & Christos, F. SocNL: Bayesian Label Propagation with Confidence. *Lecture Notes in Computer Science.* **7302**, (2012).
16. Simmel, G. *Conflict/The Web of Group Affiliations*. (Free Press, 1964).
17. Newman, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8577–82 (2006).
18. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **69**, (2004).
19. Dao, V.-L., Bothorel, C. & Lenca, P. Community structure: A comparative evaluation of community detection methods. 1–41 (2018) doi:10.1017/nws.2019.59.
20. Garza, S. E. & Schaeffer, S. E. Community detection with the Label Propagation Algorithm: A survey. *Physica A: Statistical Mechanics and its Applications* **534**, (2019).

21. Granovetter, M. Economic Action and Social Structure : The Problem of Embeddedness *American Journal of Sociology.* **91**, 481–510 (1985).
22. Granovetter, M. The Strength of Weak Ties. *American Journal of Sociology* **78**, (1973).
23. Portes, A. Social Capital: Its Origins and Applications in Modern Sociology. *Annual Review of Sociology* **24**, (1998).
24. Rivera, M. T., Soderstrom, S. B. & Uzzi, B. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology* **36**, 91–115 (2010).
25. Smith, J. A., McPherson, M. & Smith-Lovin, L. Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004. *American Sociological Review* **79**, (2014).
26. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology.* **27**, 415–444 (2009).
27. Sauder, M., Lynn, F. & Podolny, J. M. Status: Insights from Organizational Sociology. *Annual Review of Sociology* **38**, 267–283 (2012).
28. Podolny, J. M. Networks as the pipes and prisms of the market. *American Journal of Sociology* **10**, (2001).
29. Hebdige, D. *Subculture: The Meaning of Style*. (Routledge, 2012).
30. Beltagy, I., Cohan, A. & Lo, K. SciBERT: Pretrained Contextualized Embeddings for Scientific Text. *ArXiv* (2019).
31. Aceves, P. & Evans, J. A. Mobilizing Conceptual Spaces: How Word Embedding Models Can Inform Measurement and Theory Within Organization Science. *Organization Science* (2023) doi:10.1287/orsc.2023.1686.
32. Vicinanza, P., Goldberg, A. & Srivastava, S. B. A deep-learning model of prescient ideas demonstrates that they emerge from the periphery. *PNAS Nexus* **2**, (2023).
33. McPherson, J. M. & Ranger-Moore, J. R. Evolution on a dancing landscape: organizations and networks in dynamic Blau space. *Social Forces* **70**, 19–42 (1991).
34. De Domenico, M. *et al.* Mathematical formulation of multilayer networks. *Physical Review X* **3**, (2014).
35. Ma, Y., Ren, Z., Jiang, Z., Tang, J. & Yin, D. Multi-dimensional network embedding with hierarchical structure. in *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining* vols 2018-Febuary 387–395 (Association for Computing Machinery, Inc, 2018).
36. Bródka, P. & Kazienko, P. Multilayer Social Networks. in *Encyclopedia of Social Network Analysis and Mining* (2017). doi:10.1007/978-1-4614-7163-9_239-1.
37. Ma, Y. *et al. Multi-Dimensional Graph Convolutional Networks*. In *Proceedings of the 2019 siam international conference on data mining* https://epubs.siam.org/terms-privacy (2019).
38. Kipf, T. N. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907. (2016).
39. Velickovic, P. *et al.* Graph attention networks. arXiv:1710.10903v3. (2017).
40. Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. arXiv:2104.13478v2. (2021).
41. Gong, L. & Cheng, Q. *Exploiting Edge Features for Graph Neural Networks*.

42. Bacciu, D., Errica, F., Micheli, A. & Podda, M. A gentle introduction to deep learning for graphs. *Neural Networks* **129**, 203–221 (2020).
43. Zhou, Y. *et al.* Graph Neural Networks: Taxonomy, Advances, and Trends. *ACM Transactions on Intelligent Systems and Technology* **13**, (2022).
44. Burkart, N. & Huber, M. F. *A Survey on the Explainability of Supervised Machine Learning*. *Journal of Artificial Intelligence Research* vol. 70 245–317 (2021).
45. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **23**, 1–45 (2021).
46. Plebe, A. & Grasso, G. The Unbearable Shallow Understanding of Deep Learning. *Minds and Machines* **29**, 515–553 (2019).
47. Vaswani, A. *et al.* Attention is all you need. in *Advances in Neural Information Processing Systems* (2017).
48. Shalizi, C. R. & Thomas, A. C. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research* **40**, (2011).
49. McFowland, E. & Shalizi, C. R. Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association* **118**, (2023).
50. Friedkin, N. E. *A Structural Theory of Social Influence*. (Cambridge University Press, 1998).
51. Friedkin, N. E., Proskurnikov, A. V., Tempo, R. & Parsegov, S. E. Network science on belief system dynamics under logic constraints. *Science* **354**, (2016).
52. Zuckerman, E. W. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. *American Journal of Sociology* **104**, 1398–1438 (1999).
53. Buhr, H., Funk, R. J. & Owen-Smith, J. The authenticity premium: Balancing conformity and innovation in high technology industries. *Research Policy* **50**, (2021).
54. Pontikes, E. G. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations. *Administrative Science Quarterly* **57**, 81–118 (2012).
55. Hsu, G., Kocak, O. & Hannan, M. T. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. *American Sociological Review* **74**, 150–169 (2009).
56. Negro, G., Koçak, Ö. & Hsu, G. Research on categories in the sociology of organizations. *Research in the Sociology of Organizations* **31**, (2010).
57. Radio, S., Navis, C. & Glynn, M. A. How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. *Administrative Science Quarterly*. ***55,*** (2005).
58. Lopez, P., Du, C., Cohoon, J., Ram, K. & Howison, J. Mining Software Entities in Scientific Literature: Document-level NER for an Extremely Imbalance and Large-scale Task. in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* 3986–3995 (ACM, Virtual Event Queensland Australia, 2021). doi:10.1145/3459637.3481936.
59. Lane, J., Gimeno, E., Levitskaya, E., Zhang, Z. & Zigoni, A. Data Inventories for the Modern Age? Using Data Science to Open Government Data. *Harvard Data Science Review* (2022) doi:10.1162/99608f92.8a3f2336.
60. Romanelli, E. & Khessina, O. M. Regional industrial identity: Cluster configurations and economic development. *Organization Science* **16**, (2005).

61. Feldman, M. & Romanelli, E. Organizational Legacy and the Internal Dynamics of Clusters: The U.S. Human Biotherapeutics Industry, 1976–2002. in *Knowledge and Space* vol. 5 (2013).
62. Suire, R. & Vicente, J. Clusters for life or life cycles of clusters: in search of the critical factors of clusters' resilience. *Entrepreneurship and Regional Development* **26**, (2014).
63. Powell, W. W., Packalen, K. & Whittington, K. Organizational and institutional genesis the emergence of high-tech clusters in the life sciences. in *The Emergence of Organizations and Markets* (2012).
64. Dahl, M. S. & Sorenson, O. Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures. *Management Science* **58**, 1059–1071 (2012).
65. Saxenian, A. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. (Harvard University Press, 1994).
66. Bresnahan, T., Gambardella, A. & Saxenian, A. 'Old economy' inputs for 'new economy' outcomes: Cluster formation in the New Silicon Valley. *Industrial and Corporate Change* **10**, (2001).
67. Almeida, P., Dokko, G. & Rosenkopf, L. Startup Size and the Mechanisms of External Learning: Increasing Opportunity but Declining Usefulness? *Research Policy* **32**, (2003).
68. Corredoira, R. A. & Rosenkopf, L. Should auld acquaintance be forgot? The reverse transfer of knowledge through mobility ties. *Strategic Management Journal* **31**, (2010).
69. Agrawal, A. & Cockburn, I. The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International Journal of Industrial Organization* **21**, 1227–1253 (2003).
70. Feldman, M. The locational dynamics of the US biotech industry: Knowledge externalities and the anchor hypothesis. *Industry and Innovation* **10**, (2003).