Title: Accelerating science with human versus alien artificial intelligences

Authors: Jamshid Sourati[1], James A. Evans[1,2]*

Affiliations:

[1]Department of Sociology, University of Chicago; Chicago, IL, 60637, the United States.

[2]Santa Fe Institute; Santa Fe, NM, 87501, the United States.

*Corresponding author. Email: jevans@uchicago.edu

Abstract: Artificial intelligence (AI) models applied to scientific discovery typically mimic rather than complement the efforts of human scientists and inventors. Here we show that incorporating the distribution of human expertise by training unsupervised models on inferences cognitively available to experts dramatically improves AI prediction of future discoveries. These models succeed by predicting human predictions and the scientists who will make them. By tuning AI to avoid the crowd, however, we can generate scientifically promising "alien" hypotheses unlikely to be imagined or pursued without intervention until the distant future, which we demonstrate hold promise to punctuate scientific advance beyond questions currently pursued. By identifying and correcting for collective human bias, these models also suggest opportunities to improve human prediction by reformulating science education for discovery.

One-Sentence Summary: AI models for scientific discovery can accelerate advance by mimicking or avoiding the distribution of human experts.

Main Text: Research across applied science and engineering, from materials discovery to drug and vaccine development, is hampered by enormous design spaces that overwhelm researchers' ability to evaluate the full range of potentially valuable candidate designs (*1*). To face this challenge, researchers have initialized data-driven AI models with the contents of published scientific results to create powerful prediction engines that infer future findings. These models are being used to enable discovery of novel materials with desirable properties (*2*) and targeted construction of new therapies (*3*). Such efforts typically ignore the distribution of scientists, however, the human prediction engines who continuously alter the landscape of innovation. As a result, AI algorithms unwittingly compete with experts, failing to complement them and augment collective advance. As we demonstrate below, incorporating knowledge of human experts and expertise can improve predictions of future discoveries by more than 100% above content-based AI methods that ignore them. But with tens of millions of active scientists and engineers around the world, is the production of artificial intelligences that mimic human capacity our most strategic or ethical investment? By not mimicking, but rather avoiding human inferences we can design "alien" AIs that radically augment rather than replace human capacity. Compensating for the bias of collective human discovery, we demonstrate how human-avoiding or alien algorithms broaden the scope of things discovered by identifying hypotheses unlikely for scientists to imagine or pursue with amplified signals of scientific promise.

Our analysis builds on insights underlying the wisdom of crowds (*4*), which hinges on the independence and diversity of crowd members' information (*5*) and approach (*6*). In scientific crowds, findings established by more distinct methods and researchers are much more likely to replicate (*7*, *8*). If we model discovery as establishing a novel link among otherwise disconnected

concepts (*9*), no discovery can occur until discoverers arise with viewpoints diverse enough to cover the fields required to imagine connecting those concepts (Fig. 1A). This diversity of scientific viewpoints was implicitly drawn upon by Donald Swanson in a heuristic approach to knowledge generation. For example, he hypothesized that if Raynaud's disorder was linked to blood viscosity in one literature, and fish oil was known to decrease that viscosity in another, then fish oil might lessen the symptoms of Raynaud's disorder but would unlikely be arrived at by the sparse scientific community available to infer it (*10–12*), one of several hypotheses later experimentally validated (*13–15*). Our approach scales and makes this heuristic continuous, combining it with explicit measurement of the distribution of scientific expertise, and drawing upon advances in unsupervised manifold learning (*16*). Recent efforts to generate scientific hypotheses rely heavily on scientific literature, but ignore equally available publication meta-data. By programmatically incorporating information on the evolving distribution of scientific expertise, our approach balances exploration and exploitation in experimental search that enables us to both accelerate discoveries predicted to appear in future and punctuate advance by identifying promising experiments unlikely to be pursued in the near term without intervention.

The distribution of research experts across topics and time represents a critical social fact that can stably improve our inferences about whether possible scientific relationships have already or will soon be attempted. It can also inform our understanding of whether relationships will remain unimagined and unexplored until the more distant future (*9*). First, we incorporate researcher expertise and collaboration networks as useful meta-data to a precise replication of a recent, prominent analysis that predicted materials having desirable electrochemical properties from prior literature encoded with unsupervised neural network methods (*17*). We show that simply

adding this researcher information, using a formally identical approach, dramatically (~100%) improves prediction of materials to be discovered with targeted properties. Next, we extended this approach to identify a much broader matrix of materials and their functional properties (*18*), demonstrating comparable increases for predicting thousands of drugs to treat more than a hundred distinct human diseases, including vaccines and therapies for COVID-19. Finally, we use expert awareness to identify and validate the scientific and technological benefit of additionally pursuing "alien" research avenues unlikely to be explored by unassisted human experts.

Accounting for Human Experts in Machine Prediction

We model the distribution of inferences cognitively available to scientists by constructing a hypergraph over research publications. A hypergraph is a generalized graph where an edge connects a set, rather than a pair, of nodes. Our research hypergraph is mixed, containing nodes corresponding not only to materials and properties mentioned in title or abstract, but also the researchers who investigate them (Fig. 1C, first step). Random walks over this hypergraph suggest paths of inference cognitively available to active scientists, which can be used to identify mixtures of diverse expertise sufficient for discoveries. If a valuable material property (e.g., ferroelectricity—reversible electric polarization useful in sensors) is investigated by a scientist who, in prior research, worked with lead titanate ($PbTiO_3$, a ferroelectric material), that scientist is more likely to consider whether lead titanate is ferroelectric than a scientist without the research experience. If that scientist later coauthors with another who has previously worked with sodium nitrite ($NaNO_2$, another ferroelectric material), that scientist is more likely to consider that sodium nitrite may have the property through conversation than a scientist without

the personal connection. The density of random walks over this research hypergraph will be proportional to the density of cognitively plausible and conversationally attainable inferences. If two literatures share no scientists, a random walk over our hypergraph will rarely bridge them, just as a scientist will rarely consider connecting a property valued only in one with a material understood only in another (Fig. 1A).

These random walks induce meaningful proximities between nodes in our mixed hypergraph. The proximity of a material to a scientist measures the likelihood that she is or will become familiar with that concept through research experience, related reading, or social interaction. The proximity of materials to one another suggests that they may be substitutes or complements within the same experiment. The proximity of a material to a property suggests the likelihood that the material may possess the property, but also that a scientist will discover and publish it (Fig. S1A, S1B). In this way, our hypergraph-induced proximities incorporate physical and material properties latent within literature, but also the complementary distribution of scientists, enabling us to anticipate likely inferences and predict upcoming discoveries.

Our model (i) initiates a random walk over the research hypergraph with a valued property (e.g., ferroelectricity), (ii) randomly selects an article (hyperedge) with that property, (iii) randomly selects a material or author from that article, then (iv) randomly selects another article with that material or author, etc., then repeats this Markov process (*19*, *20*) (see Fig. 1B for an example). The author nodes in our hypergraph severely outnumber the materials. To compensate for this imbalance, we devise a non-uniform sampling distribution parameterized by α, which determines the fraction of material to author nodes in the resulting sequences (See Supplementary Text and Fig. S2 for details). Random walks induce similarity metrics that capture the relevance of nodes

to one another. The first metric we use draws upon the local hypergraph structure to estimate the transition probability that a random walker travels from one node to another within a fixed number of steps (see Supplementary Text). Our second metric is based on a popular, unsupervised neural network-based embedding algorithm (`deepwalk(`*21*`)`) over the generated random walks. This method is formally identical to the word embedding method used in the replicated prior work described above (*17*), but which we apply to our hypergraph, considering every random walk sequence a "sentence" linking materials, experts and functional properties (e.g., store energy; cure breast cancer, vaccinate against COVID-19). Because inferred discoveries involve relevant materials, we train the deepwalk embedding model after excluding authors from our random walk sequences (Fig. 1C). The resulting embedding maps every node to a numerical vector, with the dot-product between any pair reflecting the human-inferrable relatedness of corresponding nodes. We also create a comparable embedding space using deeper graph convolutional neural networks, which confirms the pattern of results presented here (see Methods and Supplementary Text).

Accelerating science by predicting future discoveries

To demonstrate the power of accounting for human experts, we use transition probability and deepwalk metrics to build discovery predictors and evaluate their predictions against the ground-truth discoveries that occur in reality. Moreover, we contrast our predictions with replicated word embedding-based prior work that does not account for human expertise (*17*) (labeled "word2vec") as well as a random baseline. In our numerical experiments, evaluated algorithms assess the similarity between each property and the materials available to scientists in the literature published prior to a given prediction year (e.g., 2001), then selects the 50 most

similar as predicted discoveries. Quality of predictions are evaluated based on materials discovered and published after the prediction year (see Methods for further details).

Energy-related Materials Prediction

In the first set of our experiments, we considered the valuable electrochemical properties of thermoelectricity, ferroelectricity and photovoltaic capacity against a pool of 100K candidate compounds. Following the evaluation regime of Tshitoyan et al. on the same dataset (1.5M scientific articles about inorganic materials) (*17*), we repeated identical analyses for 17 prediction periods, with prediction years ranging from 2001 to 2017, predicting future discoveries as a function of research publicly available to contemporary scientists. We computed annual precisions until the end of 2018, such that the longest precision array was nearly two decades (18 years, from 2001 to 2018) and the shortest was 2 (2017-2018, Fig. S1C). Results indicate that predictions accounting for the distribution of scientists outperformed baselines for all properties and materials by an average of 100% (Fig. 2A-C).

Drug Repurposing Prediction

We used the same approach to explore the repurposing of ~4K existing FDA-approved drugs to treat 100 important human diseases. We used the MEDLINE database of biomedical research publications and set the prediction year to 2001 (Fig. S1C). Ground-truth discoveries were based on drug-disease associations established by expert curators of the Comparative Toxicogenomics Database (CTD) (*22*), which chronicles the capacity of chemicals to influence human health. Fig. 2E reports prediction precisions 19 years after the prediction year, revealing how accounting for the distribution of biomedical experts in our unsupervised hypergraph embedding yields

predictions with 43% higher precision than identical models accounting for article content alone. We found a strong correlation between our prediction precision and drug occurrence frequency in literature ($r=0.74$, $p<0.001$), implying that our predictors work best for diseases whose relevant drugs are frequently mentioned in prior research.

COVID-19 Therapy and Vaccine Prediction

We also considered therapies and vaccines to treat or prevent SARS-CoV-2 infection. Here prediction year was set to 2020 (Fig. S1C), when the global search for relevant drugs and vaccines began in earnest. Following Gysi et al. (*23*), we considered a therapy relevant to COVID-19 if it amassed evidence to merit a COVID-related clinical trial, as reported by ClinicalTrials.gov. Results shown in Fig. 2D indicate that 36% and 38% of the predictions made by transition probability and deepwalk-based metrics entered trials within 12 months of the date of prediction (i.e., end of Dec, 2020), respectively, which further increased to 42% by the end of July, 2021. This is 350 to 400% higher than the precision of discovery candidates generated by semantic content alone (10% after the first 12 months and 12% overall). These precisions were even higher than a predictive model based on an ensemble of deep and shallow learning predictors trained on multiply measured protein interactions between COVID-19 and the pool of 3,948 relevant compounds from DrugBank (*23*), detailed and relevant information to which our model was blind (see Fig. S3 for alternative evaluation).

The success of these COVID-19 predictions suggests how fast-paced research on COVID therapies and vaccines increased the relevance of scientists' prior research experiences and relationships for the therapies and vaccines they would come to imagine, evaluate and champion

in clinical trials. Consider the clinical trial of the female progesterone for treating COVID-19 (*24*). The trial was motivated by factors including the lower global death rate of women than men from COVID-19 and anti-inflammatory properties of progesterone that may moderate the immune system's overreaction to COVID-19 in men (*25*). Random walks from our method frequently walked the path between "coronavirus" and "progesterone" literatures to predict clinical study of progesterone for coronavirus complications (Fig. S4). Our technique traced a pathway similar to the one articulated explicitly by researchers sponsoring the trial: 75% of trial-cited papers, published within the five-year period we considered in building our hypergraph (2015-2019), were identified by our prediction model, and 60% of scientists authoring those studies were sampled in our random walk sequences.

Expert-Sensitive Prediction

Our predictive models use the distribution of discovering experts to successfully improve discovery prediction. To demonstrate this, we consider the time required to make a discovery. Materials cognitively close to the community of researchers who study a given property receive greater attention and are likely to be investigated, discovered and published earlier than those further from the community. In other words, "time to discovery" should be inversely proportional to the size of the expert population aware of both property and material. We measure the size of this population by defining *expert density* as the Jaccard index of two sets of experts: those who mentioned a property and those who mentioned a specific material in recent publications. For all three electrochemical properties mentioned earlier, COVID-19 therapies and vaccines, and a majority of the 100 diseases we considered above, correlations between discovery date and expert densities were negative, significant and substantial (Fig. S5). This

confirms our hypothesis that materials considered by a larger crowd of property experts are discovered sooner. Our predictive models efficiently leverage the hypergraph of past publications to incorporate these expert densities (Fig. S6). Similar results can be derived based on embedding proximities: Fig. 3A illustrates how our predictions cluster atop density peaks in a joint embedding space of experts and the materials they investigate. These expert-material proximities are able to predict discoverers most likely to publish discoveries based on their unique research backgrounds and relationships. Moreover, computing the probability of transition from properties to expert nodes through a single intermediate material across 17 prediction years (2001 to 2017), we found that 40% of the top 50 ranked potential authors became actual discoverers of thermoelectric and ferroelectric materials one year after prediction, and 20% of the top 50 discovered novel photovoltaics (Fig. 3, bottom; see also Fig. S7).

Punctuating science by imagining the unimaginable

As illustrated above, by identifying properties and materials cognitively available to human experts, we maximize the precision of predicting published material discoveries. Our algorithm owes its success to the fact that almost all published discoveries lie in close proximity to desired properties based on our literature hypergraph (Fig. S8A). This is because scientists research and publish about materials and properties discovered through previous experience and collaborations traced by it. By contrast, if we avoid the distribution of human experts, we can produce "alien" predictions designed to complement rather than mimic the scientific community. These predictions are cognitively unavailable to human experts based on the organization of scientific fields, prevailing scientific attention, and expert education, but nevertheless manifest heightened promise for possessing desired scientific properties (Fig. S8B). Here, we propose,

pilot and test a framework that arbitrages disconnections in the hypergraph of science to identify disruptive discovery candidates more likely to possess desired properties than those that scientists investigate, which are unlikely to be discovered in the near future without machine recommendation (Fig. 1A, right).

Our framework combines two components: a *human availability* component that measures the degree to which candidate materials lie within or beyond the scope of human experts' research experiences and relationships, and a *scientific plausibility* component that amplifies predictions with promise as consistent with existing research and theory (Fig. 4A). We transform the two component scores into a unified scale and linearly combined them with a simple mixing coefficient β. Setting β=0 implies an exclusive emphasis on scientific plausibility, blind to the distribution of experts. Decreasing β imitates human experts and increasing β avoids them. At extremes, β=-1 and 1 yield algorithms that generate predictions very familiar or very strange to experts, respectively, regardless of scientific merit. Non-zero positive βs balance exploitation of relevant materials with exploration of areas unlikely considered or examined by human experts. Materials with the highest scores are reported as the algorithm's prediction and evaluated as candidates for disruptive discovery. We can quantify human availability with any graph distance metric varying with expert density (e.g., unsupervised neural embeddings, Markov transition probabilities, self-avoiding walks from Schramm-Loewner evolutions). Here we use shortest path distances between properties and materials, interlinked by authors, as above. We can quantify scientific plausibility by unsupervised embeddings of published knowledge (*17*), theory-driven simulations of material properties, or both. Here we use unsupervised knowledge embeddings for our algorithm, reserving theory-driven simulations to evaluate the value and

human complementarity of our predictions. To evaluate thermoelectric promise, power factor (PF) represents an important component of the overall thermoelectric figure of merit, zT, calculated using density functional theory for candidate materials as a strong indication of thermoelectricity (*26*, *27*). To evaluate ferroelectricity, estimates of spontaneous polarization obtained through symmetry analysis and first-principle equations serve as a reliable metric for this property (*28*). For human diseases and COVID-19, proximity between disease agents (e.g., SARS-CoV-2) and candidate compounds in protein-protein interaction networks suggests the likelihood a material will recognize and engage with the disease agent (*23*) (see Supplementary Text for more details).

We evaluate the degree to which our predictions complement human science in two ways. First, we demonstrate that more alien predictions, if discovered by the scientific community, are discovered long after prediction—much later than the scientist-mimicking predictions described earlier, once the community has filled in knowledge gaps separating valued properties and unfamiliar materials (Fig. 4B). Second, we demonstrate that more alien or human-avoiding predictions (those with a higher β) increase the scientific promise versus those made and published by the scientific community, as assessed using the precomputed theoretical scores described above. We note that this is a very conservative estimate of scientific promise for alien predictions because it is based on widely-accepted, scientist-crafted and theory-inspired or data-driven simulations. Nevertheless, even in this case, we show that by avoiding the human crowd, higher βs can improve even this conventional measure of predicted promise. Specifically, we evaluate the complementary performance of our alien predictor by verifying that the simulated quality of predictions corresponds with significantly higher values of β than for

materials actually discovered. To assess this, we computed the difference between expected values of β conditioned on when the prediction is scientifically promising and when it is found and published. A positive expectation gap measures the increase in β (or "alienness") that would maximize the prediction of materials with theoretically expected properties. As shown in Fig. 4C, for the vast majority of the properties that we considered in this section, the expectation gap is significantly positive. Building on this, we directly calculate the complementarity of a prediction from our algorithm for the human scientific community in two ways. First, we compute the probability that it is plausible but not discoverable for any particular value of β (Figs. S8D, S8E). Second, we note the minimal overlap between published discoveries and alien predictions, which nevertheless retain or increase their theoretical merit (Fig. S9).

Discussion

We demonstrate the power of incorporating expert-awareness into artificial intelligence systems for accelerating and punctuating future discovery. Our models succeed by directly predicting human discoveries and the human experts who will make them, yielding an average of 100% improvement in prediction precision. By tuning these algorithms to avoid the crowd, however, they generate even more promising hypotheses unlikely to be imagined, pursued or published without machine recommendation for years into the future. By identifying and correcting for collective patterns of human attention, formed by field boundaries and institutionalized education, these models complement the contemporary scientific community. A further class of alien predictions could be tuned to compensate not only for emergent collective bias, but universal cognitive constraints, such as limits on the human capacity to conceive or search through complex combinations (e.g., high-order therapeutic cocktails (*29*)). Disorienting

hypotheses from such a system will not be beautiful, but being inconceivable, they break fresh ground and sidestep the path-dependent "burden of knowledge" where scientific institutions require new advances built upon the old for ratification and support (*30*, *31*).

Our approach can also be used to identify individual and collective biases that limit productive exploration and suggest opportunities to improve human prediction by reformulating science education for discovery. Insofar as research experiences and relationships condition the questions scientists investigate, education tuned to discovery would conceive of each student as a new experiment, recombining knowledge and opportunity in novel ways. Our investigation underscores the power of incorporating human and social factors to produce artificial intelligence that complements rather than substitutes for human expertise. By making AI hypothesis generation aware of human expertise, it can race with rather than against the scientific community to expand the scope of human imagination and discovery.

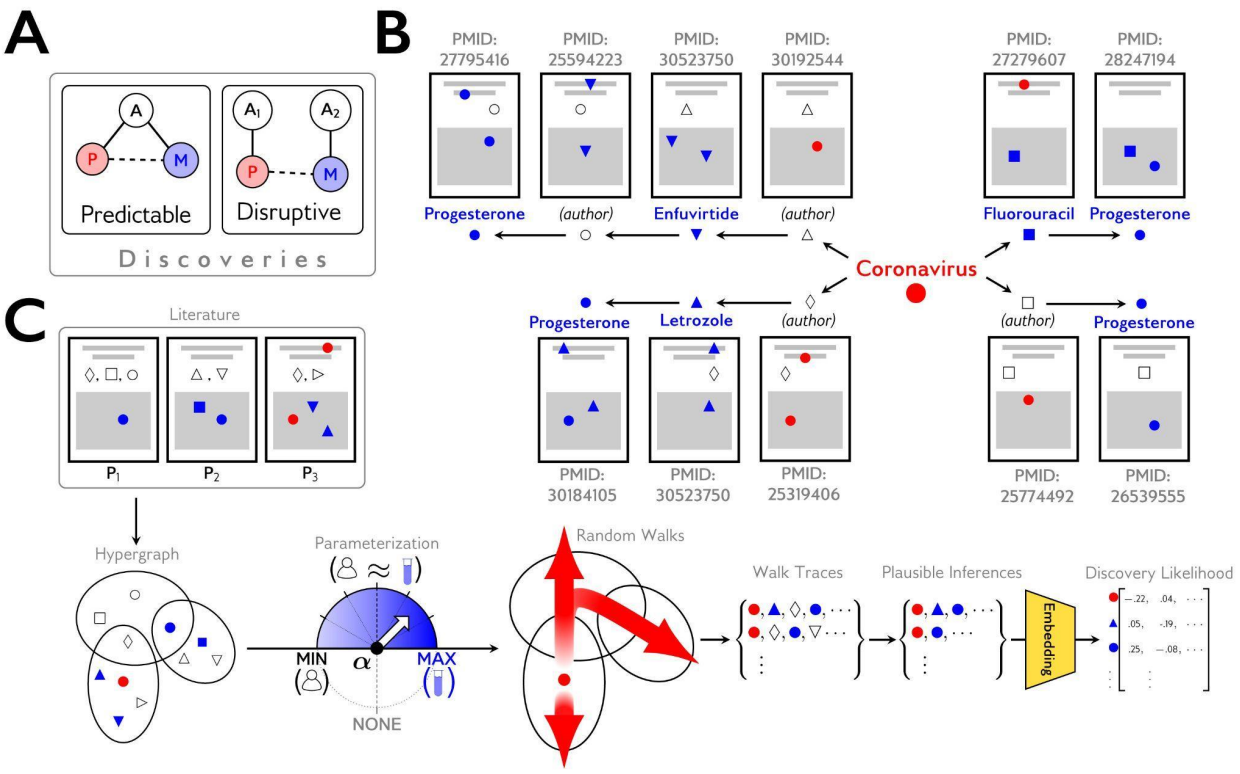

Fig. 1. (A) Two possible scenarios when there exists a hidden underlying relationship between material M and property P waiting to be discovered. Uncolored circles represent non-overlapping populations of human experts and colored nodes indicate a material (colored in blue) or a property (colored in red). Solid lines between uncolored and colored nodes imply that the experts represented by the former studied or have experience with the material or property denoted by the latter. Dashed lines represent existing property-material links that have not been discovered yet. The P-M relation in the left scenario is likely to be discovered and published in the near future, but is likely to escape scientists' attention in the right scenario; it would disrupt the current course of science. **(B)** Four example random walk paths starting from "Coronavirus" node (property) and ending in "Progesterone" (a chemical under clinical trial investigation for therapeutic efficacy). Each arrow connecting two nodes indicates a sampling step, where the paper shown on top of the receiving node comprises a hyperedge containing both the prior step and the current step sampled nodes. **(C)** Illustration of our hypergraph deepwalk algorithm: 1) We construct a hypergraph based on a literature represented by three papers here. Uncolored shapes represent authors and colored shapes indicate properties (red) or materials (blue) mentioned in the title and abstract of articles. 2) We perform classic or α-modified random walk sampling, which 3) results in a set of sequences consisting of authors, materials and the considered property. 4) We remove authors from the sequences to keep only the materials on which the discovery inference will be applied. 5) We train a word embedding model

(e.g., Word2Vec) on the sampled sequences of material/property tokens, which results in 6) a vector representation of the materials and property we use to compute pairwise cosine similarities for prediction.

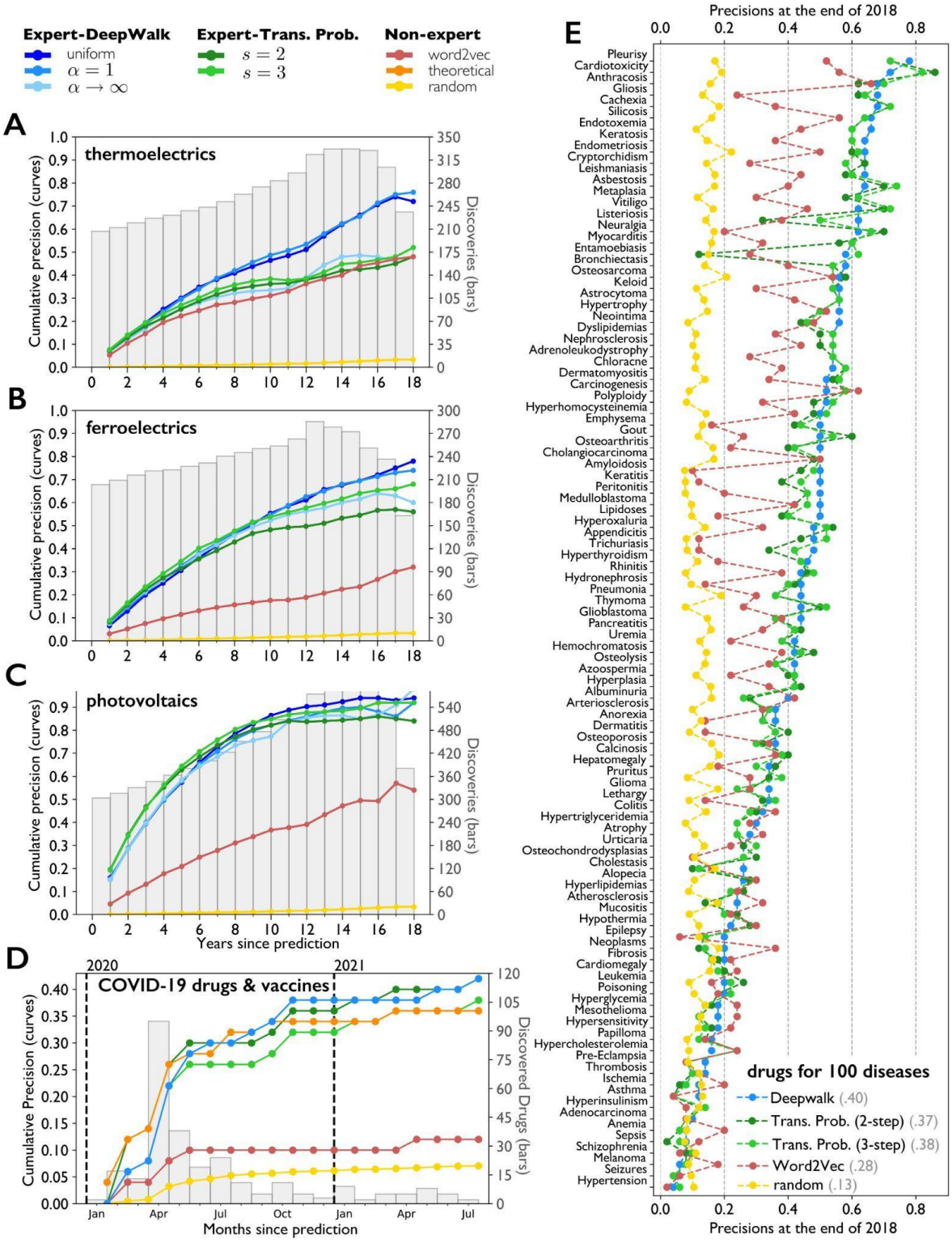

Fig. 2. Precision rates of discovery for materials associated with different properties and prediction years: **(A-C)** chemical compounds and electrochemical properties including thermoelectricity, ferroelectricity, and photovoltaic capacity, respectively,

with prediction years varying from 2001 to 2017; **(D)** therapeutics and vaccines for COVID-19 for the prediction year 2020; **(E)** general disease-drug associations for prediction year 2001. Precisions reported for general disease-drug associations are individual rates computed 19 years after prediction year, but computed annually for electrochemical properties and monthly for COVID-19 efficacy (See Fig. S1C). Gray bars in Figs. **(A-D)** indicate the number of new discoveries occurring in reality for each month or year of the prediction period. Predictions accounting for the scientific community are far superior to those that ignore them.

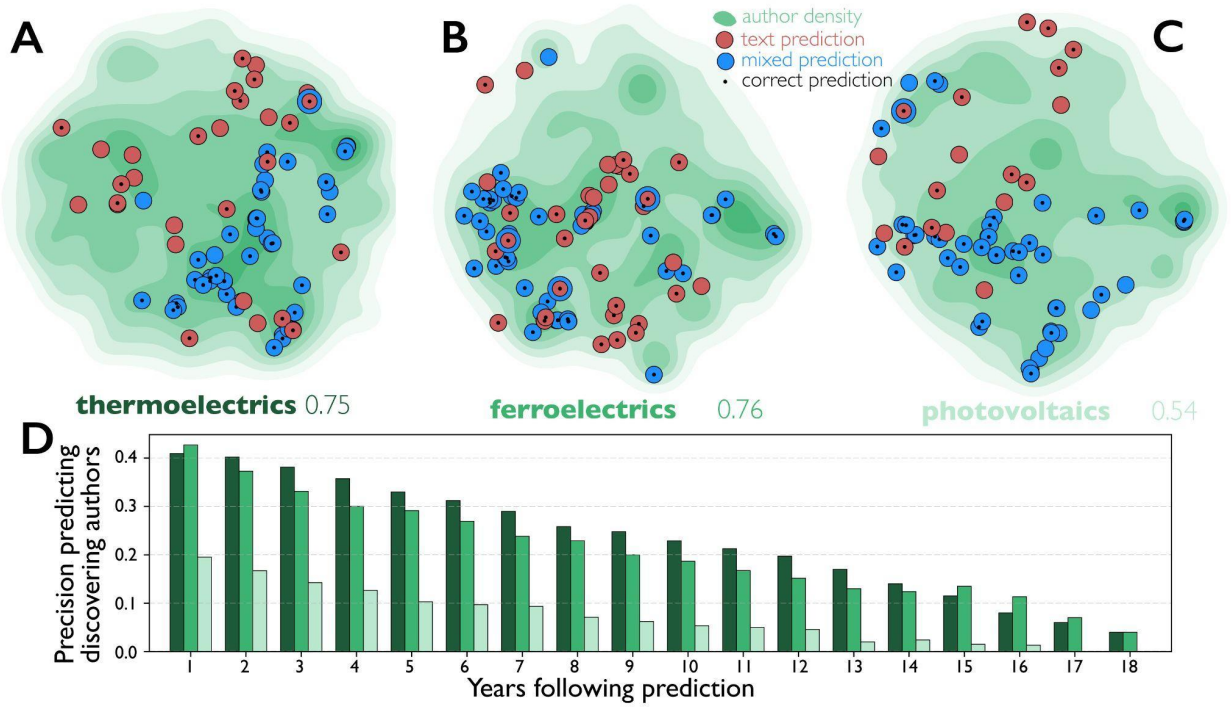

Fig. 3. (A-C) 2D projections of expert-sensitive material predictions made by deepwalk (blue circles) and content-exclusive word2vec model (red circles) for thermoelectricity (left), ferroelectricity (center) and photovoltaic capacity (right). Circles with center dots indicate true positive predictions discovered and published in subsequent years and empty circles are false positives. Predictions are plotted atop the density of experts (topo map and contours estimated by Kernel Density Estimation) in a 2D tSNE-projected embedding space. Before applying tSNE dimensionality reduction, the original embedding was obtained by training a word2vec model over sampled random walks across the hypergraph of published science. Red circles are more uniformly distributed, but blue circles concentrate near peaks of expert density. **(D)** Precision rates for predicting discoverers of materials with electrochemical properties. Predictive models are built based on two-step transitions between property and expert nodes with an intermediate material in the transition path. Bars show average precision of expert predictions for individual years. An expert can publish a discovery in multiple years. Total precision rates are also shown near each property ignoring the repetition of discovering experts.

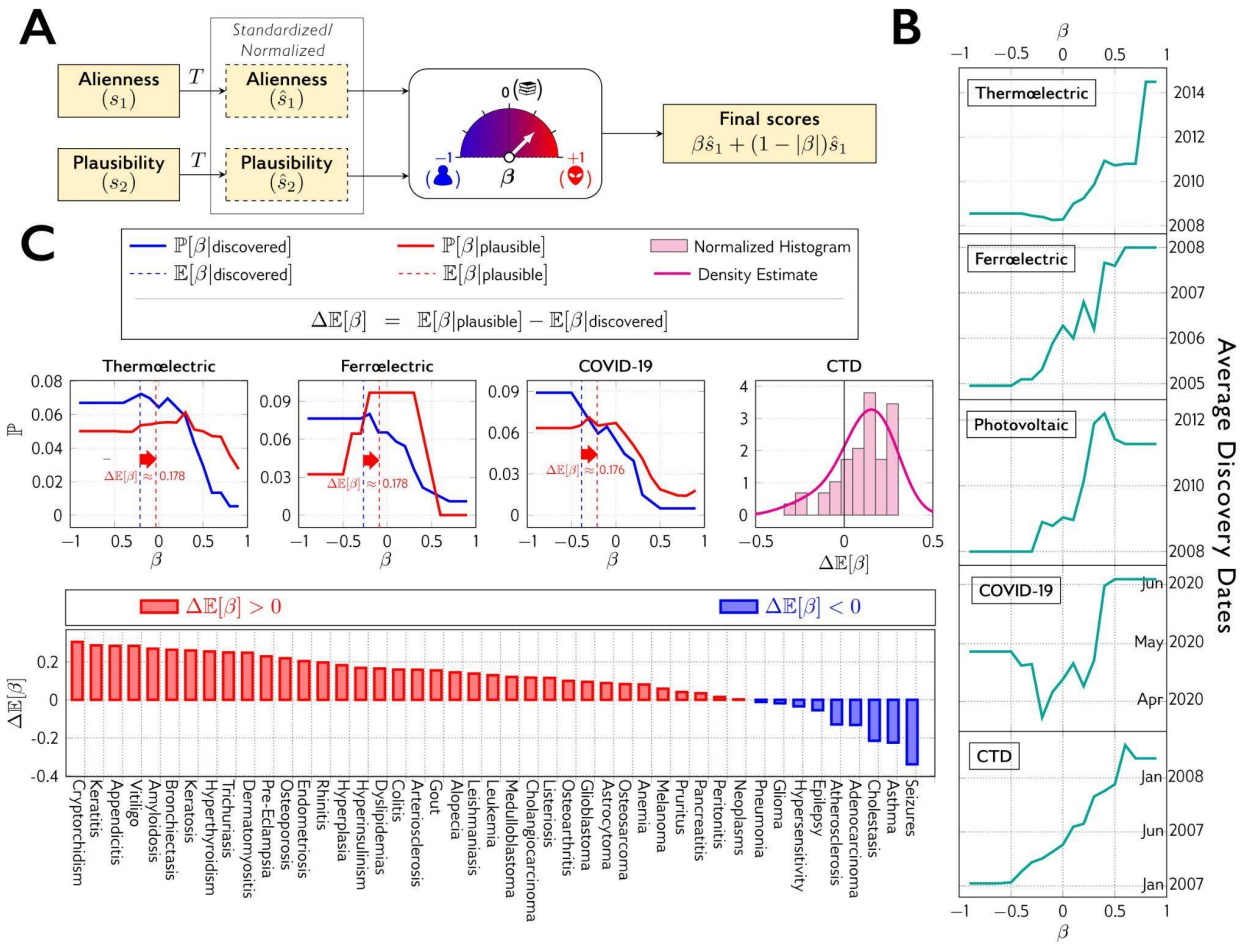

Fig. 4. (A) Diagram of our alien artificial intelligence. In the first step, the scores undergo the same transformation T, which is a combination of standardization and normalization, to be mapped into comparable scales. The transformed scores will then be linearly combined with parameterized by signed coefficient β varying from -1 (the most human-like prediction) to +1 (the most alien prediction) via 0 (neutral predictions, blind to human availability). For intermediate values of β, both plausibility and alienness contribute to the final scores. When β>0, alienness contributes positively and when β<0, its contribution is negative (encouraging human-like predictions). **(B)** Average discovery dates for the discovered predictions generated with various values of β. **(C)** Plots of β likelihoods given plausibility and discoverability when varying β from -1 to +1, and the resulting expectation gaps for each property. In the case of CTD diseases, where we only considered 45 diseases for which we had available theoretical drug relevance scores (see Supplementary Text), the final gap has been shown individually (bar chart) and collectively (the histogram).

References

1. S. R. Khadherbhi, K. S. Babu, Big Data Search Space Reduction Based On User Perspective Using Map Reduce. *International Journal of Advanced Technology and Innovative Research*. **7**, 3642–3647 (2015).

2. B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering. *Science*. **361**, 360–365 (2018).

3. E. Smalley, AI-powered drug discovery captures pharma interest. *Nat. Biotechnol.* **35**, 604–605 (2017).

4. F. Galton, Vox populi (the wisdom of crowds). *Nature*. **75**, 450–451 (1907).

5. J. Surowiecki, The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*. **296** (2004).

6. S. E. Page, *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy* (Princeton University Press, 2019).

7. V. Danchev, A. Rzhetsky, J. A. Evans, Centralized scientific communities are less likely to generate replicable results. *Elife*. **8** (2019), doi:10.7554/eLife.43094.

8. A. V. Belikov, A. Rzhetsky, J. Evans, Detecting signal from science:The structure of research communities and prior knowledge improves prediction of genetic regulatory experiments. *arXiv [cs.SI]* (2020), (available at http://arxiv.org/abs/2008.09985).

9. A. Rzhetsky, J. G. Foster, I. T. Foster, J. A. Evans, Choosing experiments to accelerate collective discovery. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14569–14574 (2015).

10. D. R. Swanson, Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect. Biol. Med.* **30**, 7–18 (1986).

11. D. R. Swanson, Medical literature as a potential source of new knowledge. *Bull. Med. Libr. Assoc.* **78**, 29–37 (1990).

12. M. Weeber, H. Klein, L. T. W. de Jong-van den Berg, R. Vos, Using concepts in literature-based discovery: Simulating Swanson's Raynaud--fish oil and migraine--magnesium discoveries. *J. Am. Soc. Inf. Sci. Technol.* **52**, 548–557 (2001).

13. J. Evans, A. Rzhetsky, Machine Science. *Science*. **329**, 399–400 (2010).

14. R. A. Digiacomo, J. M. Kremer, D. M. Shah, Fish-oil dietary supplementation in patients with Raynaud's phenomenon: A double-blind, controlled, prospective study. *The American*

Journal of Medicine. **86** (1989), pp. 158–164.

15. H.-Y. Chiu, T.-H. Yeh, Y.-C. Huang, P.-Y. Chen, Effects of Intravenous and Oral Magnesium on Reducing Migraine: A Meta-analysis of Randomized Controlled Trials. *Pain Physician*. **19**, E97–112 (2016).

16. T. Mikolov, W. Yih, G. Zweig, Linguistic regularities in continuous space word representations. *hlt-Naacl* (2013) (available at http://www.aclweb.org/anthology/N13-1#page=784).

17. V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature*. **571**, 95–98 (2019).

18. B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, A. I. Cooper, A mobile robotic chemist. *Nature*. **583**, 237–241 (2020).

19. F. Shi, J. G. Foster, J. A. Evans, Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Soc. Networks*. **43**, 73–85 (2015).

20. U. Chitra, B. Raphael, in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri, R. Salakhutdinov, Eds. (PMLR, 2019), vol. 97 of *Proceedings of Machine Learning Research*, pp. 1172–1181.

21. B. Perozzi, R. Al-Rfou, S. Skiena, in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (Association for Computing Machinery, New York, NY, USA, 2014), *KDD '14*, pp. 701–710.

22. A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, R. McMorran, J. Wiegers, T. C. Wiegers, C. J. Mattingly, The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).

23. D. M. Gysi, Í. Do Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R. M. Baron, D. Ghiassian, J. Loscalzo, A.-L. Barabási, Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19. *ArXiv* (2020) (available at https://www.ncbi.nlm.nih.gov/pubmed/32550253).

24. S. Ghandehari, Progesterone for the Treatment of COVID-19 in Hospitalized Men. *Identifier NCT04365127* (2020), (available at https://clinicaltrials.gov/ct2/show/study/NCT04365127).

25. S. Ghandehari, Y. Matusov, S. Pepkowitz, D. Stein, T. Kaderi, D. Narayanan, J. Hwang, S. Chang, R. Goodman, H. Ghandehari, J. Mirocha, C. Bresee, V. Tapson, M. Lewis, Progesterone in Addition to Standard of Care Versus Standard of Care Alone in the Treatment of Men Hospitalized with Moderate to Severe COVID-19: A Randomized, Controlled Pilot Trial. *Chest* (2021), doi:10.1016/j.chest.2021.02.024.

26. A. Mehdizadeh Dehkordi, M. Zebarjadi, J. He, T. M. Tritt, Thermoelectric power factor: Enhancement mechanisms and strategies for higher performance thermoelectric materials. *Mater. Sci. Eng. R Rep.* **97**, 1–22 (2015).

27. F. Ricci, W. Chen, U. Aydemir, G. J. Snyder, G.-M. Rignanese, A. Jain, G. Hautier, An ab initio electronic transport database for inorganic materials. *Sci Data*. **4**, 170085 (2017).

28. T. E. Smidt, S. A. Mack, S. E. Reyes-Lillo, A. Jain, J. B. Neaton, An automatically curated first-principles database of ferroelectrics. *Sci Data*. **7**, 72 (2020).

29. L. K. Gediya, V. C. Njar, Promise and challenges in drug discovery and development of hybrid anticancer drugs. *Expert Opin. Drug Discov.* **4**, 1099–1111 (2009).

30. B. F. Jones, The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).

31. M. Szell, Y. Ma, R. Sinatra, A Nobel opportunity for interdisciplinarity. *Nat. Phys.* **14**, 1075–1078 (2018).

Acknowledgments: The authors wish to thank our funders for their generous support: National Science Foundation #1829366; Air Force Office of Scientific Research #FA9550-19-1-0354, #FA9550-15-1-0162; DARPA #HR00111820006. We thank Laszlo Barabasi and Deisy Morselli Gysi for helpful data related to their network-based forecast of COVID-19 drugs and vaccines with protein-protein interactions (*23*), and Anubhav Jain, Vahe Tshitoyan and Alex Dunn for sharing data and code to help replicate their work on unsupervised word embeddings and latent knowledge about material science(*17*). We also thank participants of the Santa Fe Institute workshop "Foundations of Intelligence in Natural and Artificial Systems", the University of Wisconsin at Madison's HAMLET workshop, and colleagues at the Knowledge Lab for helpful comments.

Funding:

National Science Foundation grant 1829366 (JE)

Air Force Office of Scientific Research grant FA9550-19-1-0354 (JS, JE)

Air Force Office of Scientific Research grant FA9550-15-1-0162 (JS, JE)

Defense Advanced Research Projects Agency grant HR00111820006 (JE)

Author contributions: Each author's contribution(s) to the paper should be listed [we encourage you to follow the CRediT model]. Each CRediT role should have its own line, and there should not be any punctuation in the initials.

Examples:

Conceptualization: JE

Data curation: JS

Formal analysis: JE, JS

Methodology: JS, JE

Software: JS

Visualization: JS, JE

Funding acquisition: JE

Project administration: JE, JS

Writing – original draft: JS

Writing – review & editing: JE

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data, code, and materials used in the analysis are available through the following link: https://github.com/jsourati/social-knowledge-analysis.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S9

Fig. 1. How and why our expert-aware discovery prediction algorithm works, illustrated with an example. (A) Two possible scenarios when there exists a hidden underlying relationship between material M and property P waiting to be discovered. Uncolored circles represent non-overlapping populations of human experts and colored nodes indicate a material (colored in blue) or a property (colored in red). Solid lines between uncolored and colored nodes imply that the experts represented by the former studied or have experience with the material or property denoted by the latter. Dashed lines represent existing property-material links that have not been discovered yet. The P-M relation in the left scenario is likely to be discovered and published in the near future, but is likely to escape scientists' attention in the right scenario; it would disrupt the current course of science. (B) Four example random walk paths starting from "Coronavirus" node (property) and ending in "Progesterone" (a chemical under clinical trial investigation for therapeutic efficacy). Each arrow connecting two nodes indicates a sampling step, where the paper shown on top of the receiving node comprises a hyperedge containing both the prior step and the current step sampled nodes. (C) Illustration of our hypergraph deepwalk algorithm: 1) We construct a hypergraph based on a literature represented by three papers here. Uncolored

shapes represent authors and colored shapes indicate properties (red) or materials (blue) mentioned in the title and abstract of articles. 2) We perform classic or α-modified random walk sampling, which 3) results in a set of sequences consisting of authors, materials and the considered property. 4) We remove authors from the sequences to keep only the materials on which the discovery inference will be applied. 5) We train a word embedding model (e.g., Word2Vec) on the sampled sequences of material/property tokens, which results in 6) a vector representation of the materials and property we use to compute pairwise cosine similarities for prediction.

Fig. 2. Evaluating our expert-aware discovery predictions against various baselines. Precision rates of discovery for materials associated with different properties and prediction years: (A-C) chemical compounds and electrochemical properties including thermoelectricity, ferroelectricity, and photovoltaic capacity, respectively, with prediction years varying from 2001 to 2017; (D) therapeutics and vaccines for COVID-19 for the prediction year 2020; (E) general disease-drug associations for prediction year 2001. Precisions reported for general disease-drug associations are individual rates computed 19 years after prediction year, but computed annually for electrochemical properties and monthly for COVID-19 efficacy. Gray bars in Figs. (A-D) indicate the number of new discoveries occurring in reality for each month or year of the prediction period. Predictions accounting for the scientific community are far superior to those that ignore them.

Fig. 3. Visualizing the distribution of authors in the vicinity of our expert-aware predictions versus content-exclusive predictions, and illustrating the performance of our hypyergraph-based algorithm in identifying the discovering authors. (A-C) 2D projections of expert-sensitive material predictions made by deepwalk (blue circles) and content-exclusive word2vec model (red circles) for thermoelectricity (left), ferroelectricity (center) and photovoltaic capacity (right). Circles with center dots indicate true positive predictions discovered and published in subsequent years and empty circles are false positives. Predictions are plotted atop the density of experts (topo map and contours estimated by Kernel Density Estimation) in a 2D tSNE-projected embedding space. Before applying tSNE dimensionality reduction, the original embedding was obtained by training a word2vec model over sampled random walks across the hypergraph of published science. Red circles are more uniformly distributed, but blue circles concentrate near peaks of expert density. (D) Precision rates for predicting discoverers of materials with electrochemical properties. Predictive models are built based on two-step transitions between property and expert nodes with an intermediate material in the transition path. Bars show average precision of expert predictions for individual years. An expert can publish a discovery in multiple years. Total precision rates are also shown near each property ignoring the repetition of discovering experts.

Fig. 4. Design and performance of "alien" AI predictions. (A) Diagram of our alien artificial intelligence. In the first step, the scores undergo the same transformation T, which is a combination of standardization and normalization, to be mapped into comparable scales. The transformed scores will then be linearly combined with parameterized by signed coefficient β varying from -1 (the most human-like prediction) to +1 (the most alien prediction) via 0 (neutral predictions, blind to human availability). For intermediate values of β, both plausibility and alienness contribute to the final scores. When $\beta>0$, alienness contributes positively and when

β<0, its contribution is negative (encouraging human-like predictions). (B) Average discovery dates for the discovered predictions generated with various values of β. (C) Plots of β likelihoods given plausibility and discoverability when varying β from -1 to +1, and the resulting expectation gaps for each property. In the case of CTD diseases, where we only considered 45 diseases for which we had available theoretical drug relevance scores (see Supplementary Text), the final gap has been shown individually (bar chart) and collectively (the histogram).