Notes on 100x (vs <2x) Research Opportunities

Ravi B. Sojitra
Stanford University

First draft: April 9, 2025
This draft: May 9, 2025

Summary

- Last week, I attended two Stanford Artificial Intelligence (AI) workshops: one on AI Alignment for Businesses and one on AI and Experiment Design for Bioengineering.
- My goal was to understand which applied statistics, causal inference, and machine learning contributions would accelerate the pace of innovation and scale of impact.
- I documented opportunities based on the workshops and my own experiences: (1) automatic, scalable custom measurement; (2) better AI agents; (3) technology to support compliance and privacy requirements; (4) patents and regulatory frameworks.
- Submit feedback by creating a GitHub Issue at the `public-notes` repository. Thanks!

Motivation

Eleven years ago, I shifted my focus from biology and neuroscience to applied statistics, causal inference, and machine learning. I did this because I love the latter's pace of innovation and real world, world-wide impact via industrial adoption. I also enjoy the work.

One recent example is the impact of generative models on biology research and development (R&D). More specifically, de novo drug design, protein folding prediction, protein synthesis, genomics, and transcription. At a Stanford workshop on AI and Experiment Design for Bioengineering, I learned that R&D in machine learning (e.g. diffusion models), data curation efforts (e.g. the Protein Data Bank), and computing (e.g. software, hardware) *already* unlocked hundreds of product categories, technological capabilities, and regulatory challenges. I have a feeling that I am understating the impact potential.

Question: How can we bring the future to the present, faster?

I attend workshops to identify methodological research directions that could accelerate innovation. Last week, I focused on Agentic AI products and bioengineering at workshops on AI Alignment for Businesses and AI and Experiment Design for Bioengineering.

Four Positive Answers (10-100x Opportunities)[1]

1. Automate and Scale *Custom* Data Collection and Experiment Design

One set of pain points that resonated with me during both workshops is measurement challenges. For example, at the AI Alignment for Business workshop, measuring model performance across many languages and cultural norms was one business challenge. At the AI & Experiment Design for Bioengineering workshop, experts reminded the audience that technological limitations in measuring structures, reactions, and cell states have historically been the limiting factors. These points resonated with me because I experienced frustrations with measurement limitations across my academic work (e.g. explainable AI, neuroscience, psychology) and industrial work (search, marketing, ranking, targeting). For each project, I spent months or years diagnosing data and experiment issues, collecting different types of data after failed experiments, and running new experiments with improved interventions only to understand tiny slices of larger pictures. When you have to collect data yourself from the real world, each project iteration takes a long time.

One recent, small success in this direction is related to detecting and measuring content policy violations on digital platforms. Companies are consistently finding that Multimodal Large Language Models can be cleverly used to improve measurement and detection of content policy violations. Violations that were previously difficult to detect and label are now being caught at scale. This not only improves the products, but enables better research about the content ecosystem and users.

Nevertheless, measurement remains a major friction to innovation even at well resourced internet companies. For example, content ranking teams frequently run experiments to evaluate ranking algorithm changes. If we suspect that a treatment arm in an experiment had a negative impact on user experience even though engagement and revenue increased, we should check other metrics. For example, we should check whether there was an increase in content reported by users or deleted content. However, if this information is not logged ahead of time, a nontrivial amount of engineering investment and statistics expertise is required to log the necessary data before the experiment can be re-run. We have to make sure it is done correctly and with imperceptible latency for users.

One consistent pattern across emprical settings is that when empiricists learn more about mechanisms in an experiment, the next steps are to measure new things with greater precision (e.g. lifestyle choices, biological or molecular structures, browsing habits). Methods for automatically designing cheap measurement tools (e.g. imaging or data logging) given a scientist's needs would be game changers. In theory, "cloud laboratories" could fulfill this function, but the current quality and scale of services are unsatisfactory.

[1] **Disclaimer**. This is not a summary. I am synthesizing learnings with prior experiences as an exercise to improve future time investments. My perspectives evolve over time.

2. Build Better AI Agents for Tasks that Distract from Intellectual Work

I have noticed three qualitatively different approaches to Agentic AI development. One approach is to build AI agent systems that autonomously generate hypotheses based on existing literature, design experiments, analyze data, and propose new hypotheses based on results for the user (e.g. scientist) to verify or investigate further. A second approach is to build a system that executes conceptual subroutines in a workflow so that the user can focus on less conceptually algorithmic tasks. A third approach extends the second by giving the AI system access to tools. There are merits and limitations to all of them.

One illustration of the first approach can be found in Huang et al. (2025). The authors attempted to automate scientific discovery. Authors of such approaches are typically transparent about the limitations. In particular, current agentic AI systems (for research) only occasionally propose novel and plausible scientific hypotheses that are worth investigating further. However, because these systems propose ideas so quickly (within minutes or hours), they could still accelerate innovation if the yield can be sustained in the longer term. Right now, this is unlikely, but it seems possible to achieve this in a couple years for very narrow sets or types of questions.

A more modest, and currently more reliable, approach to developing usable AI agents is to translate conceptual subroutines in workflows into prompts that AI agents tend to respond to correctly. For example, AhmadiTeshnizi et al. (2023) propose a solution for automatically formulating mathematical optimization problems in a way that state-of-the-art software can be applied. Although such agents are not yet as reliable as experts, they reduce the barriers and can save time.

Third, taking things a step further, the second approach works astonishingly well when AI agents are trained to use tools (e.g. calculators). I recently met a startup founder whose AI solution easily achieved 100% accuracy on tasks like math calculations by using well designed prompts and allowing the AI system to use tools to execute tasks. I do not want to free marketing since they advertised their results in a slide deck, and I have no way of assessing quality of their evaluation. However, if you want to know who, feel free to reach out to me.

I feel confident now that the necessary building blocks already exist for automating workflows that are conceptually iterative or algorithmic in nature, but not defined well enough a priori to write software. For example, if your typical biology project workflow for understanding a mechanism is to do a literature review, knock out some genes, take measurements after some time, and propose alternative genes to knock out based on the results, a lot of this work can and should be automated, and the time savings should be reinvested in thinking harder about which mechanisms are worth understanding.

I am looking forward to future of Agentic AI solutions. It does not make sense to me that we invest so much time and money in training scientists and engineers only to have them perform administrative tasks, diagnose contamination issues in experiments, and fix coding or math bugs. We should be striving to reallocate time from such tasks to planning more ambitious research designs and asking better questions.

3. Plan for and Shape the Future of Compliance and Privacy Engineering

I learned a lot of new things from the workshops, and I documented some of these lessons in the last few pages. The most important lesson for me is that there is no clear industry standard (yet) on how to approach compliance and privacy engineering for Agentic AI products. The main paying customers are businesses, so meeting their compliance and privacy needs is necessary for generating revenue. This theme is consistent with my impression about machine learning and AI products in general. Moving forward, I expect methodological contributions that raise fewer red flags during legal, policy, and privacy reviews will have faster and broader adoption.

4. Shape the Future of Patents and Regulatory Frameworks

Experts noted that patent offices are not prepared for emerging challenges. For example, since it has become easier to synthesize artificial proteins, it has become easier to copy patented molecules while being "different enough". In the past, you could not write down a shopping list of properties you want, and then generate a protein that has all of those properties. Now, this is increasingly possible, and it raises at least two challenges.

First, this increases incentives for mimicing commercially successful technologies, thereby reducing incentives for R&D investment in problems without working solutions. For example, biotechnologies are so resource intensive during drug discovery, development, and evaluation that large, long term financial incentives (e.g. through intellectual property rights) are necessary to attract investment. My understanding is that (1) artificial human gene sequences and (2) amino acid sequences are patentable if they meet a few requirements. If the future of engineering these sequences is increasingly programmatic, it will be cheaper to mimic commercially successful products. While this would probably reduce the prices of drugs and technologies that are expensive today, this may reduce investment and innovation in areas without working solutions. I want more of the latter.

Second, it sounds like there is a strong financial incentive to preemptively patent many artificial amino acid and gene sequences by leveraging improvements in predictive modeling (e.g. protein folding). Imagine having exclusive access to a machine that accurately predicts structures of proteins very different from the ones we know of today (i.e. a machine better than AlphaFold). This would enable you to design proteins for unsolved bioengineering challenges and patent large sets of artificial gene and amino acid sequences. It is hard for me to believe that investors can resist the idea of patenting the most promising biological and chemical products and licensing them for revenue later on.

In the latter case, the return on investment would be tied more closely to the quality of scientific evaluation compared to startups we are used to seeing. As a scientist, it is exciting to imagine a future where scientific discovery is the core business function, but I am worried that the current thinking about patents has put us on the wrong path. There is already public discourse about artwork and copyright, but I have heard little about novel physical product categories.

For example, should we introduce algorithmic or quantitative definitions of novelty and ownership to support qualitative desiderata? Is that even a sensible way to think about things in a world where bioengineering is more programmatic? Scientists need to think hard about these questions and prepare good recommendations to inform regulatory frameworks.

Two Negative Answers ($<$ 2x Opportunities)

One thing that stands out to me is that many objectives methods researchers focus on have limited impact opportunity. Even in the best case scenario. While the examples below are important, the opportunities in the previous section are orders of magnitudes larger.

1. Increase Statistical Power ("Sample efficiency")

This is a favorite topic among methods researchers. For example, algorithms from active learning and Bayesian optimization research are tried and true methods for cost cutting because less resources have to be allocated to learning hyperparameters. Applications include materials science and hyperparameter tuning at technology companies. However, the impact of ***new*** methods on the pace of innovation appears to be qualitatively smaller than the examples I shared in the "Positive Answers' section. This suggests to me that other blockers and frictions are the limiting factors. For example, academic and industrial norms or underinvestment in planning quality research designs.

I suspect exceptions to this rule are contributions that are valuable for reasons others than sample efficiency. For example, Dang et al. (2025) propose a system for scaling drug discovery that leverages domain expertise by using a tool to collect expert preferences. Although it improves statistical efficiency, I think such interfaces for data collection have much greater potential as tools used by an Agentic AI system.

2. Raise the (Publication) Bar

One discussion point was whether the research publication bar ought to be raised. In particular, the question was whether randomized experiment results should be required for publication. The rationale is that the volume of papers is unmanageable and accelerating in growth, but most of it is underwhelming, unimpactful, and dishonest about the facts. Nevertheless, the consensus is usually that these efforts will slow the pace of innovation.

Some Notes on AI Alignment for Businesses

Compliance with regulation and company policies was a major theme. Providing a commercial Agentic AI product is challenging when you need to control who is allowed access to what information. For example, you may not want your performance evaluation or social security number available to everyone who uses the company's Agentic AI tools.

Alignment was defined more precisely than I have seen in engineering and computer science departments. Not surprisingly, in practice, this is defined roughly as "alignment with respect to the company's governance structure".

Example desiderata of a generative AI based search or agent system.

- You want the Agentic AI system to respect governance structures. For example, decision making permissions and information access for different groups of employees.
- You may not want to disclose information to people outside your reporting chain.
- You want to be aware of the documents and people you are elevating as experts when providing AI search results.
- You want to minimize political bias in internal AI systems.

Technical challenges in Agentic AI systems.

- How should you ingest data from various sources (e.g. Google Docs, Microsoft Word, Overleaf, text files, PDFs, data lakes, other tools)?
- How should you create datasets from conversations between employees (e.g. discussions between networks of people on Slack across different chat channels)?

Some Notes on AI & Experiment Design for Bioengineering

The conference program highlighted work on AI agents, Active Learning, Bayesian Optimization, Diffusion Models, and Transformer based Sequence Models. Other topics were discussed, but not emphasized. For example, there were also presentations on mathematical optimization and conformal inference. There was surprisingly little discussion about compliance, regulation, and property rights.

Overall, there was a lot of excitement about generative models. Diffusion models have more attention, empirical evidence, and impact than I expected. Conditional sampling using diffusion models modeling stood out to me in particular. I am not sure to what extent this excitement is due to audience selection bias, but the empirics are impressive.

Empirical Observations Noted by Experts

- The Reinforcement Learning From Human Feedback (RLHF) paradigm works when combined with data from experiments testing which structures repel and bind to one another (potentially in the presence of other structures).
- Genotype to phenotype mapping is a problem where the former has a lot of cheaper data and the latter has little data.
- Cell state trajectory can be represented in low dimensions using cytometry data in silico.