

AI, Labor, and Economy Case Studies In-Brief

# Zymergen



**PARTNERSHIP ON AI**

Zymergen is a biotech startup in the Bay Area, California, that uses machine learning to improve scientific experiment design and employs high-throughput screening in its 'automated wet lab' to execute experiments much faster than a conventional lab.<sup>1</sup> As an "AI-native"<sup>2</sup> company, Zymergen provides unique insights into the economic, organizational, and labor implications for companies that are designed at their conception around the use of artificial intelligence.

## Case background

Zymergen was founded in 2013 with the core principle of using AI-related technologies as a key differentiator. The company has raised \$574M from investors such as SoftBank Vision Fund, Data Collective, ICONIQ Capital, and McKinsey & Company (a co-author of the case studies), among others.<sup>3</sup> Zymergen provides R&D services to its customers, aiming to deliver improved economics for fermentation products in sectors such as agricultural commodities and electronics through specific microbial strains. Zymergen also leverages its insights and platform to develop its own products for the electronics (films, coatings, adhesives), marine, and personal care industries. The global market for fermentation products was roughly \$150 billion in 2016, yet this capital-intensive set of industries is highly competitive, most with thin operating margins.<sup>4</sup> Manufacturers often use internal R&D labs or external R&D vendors, such as Zymergen, to improve the economics of microbial strains for large-scale manufacturing or to avoid having to invest in new production capacity.

Typically, in microbial strain improvements, R&D research teams take hypothesis-driven approaches by building on the present state of scientific knowledge and formulating an experiment design that they believe could improve a particular strain. Zymergen differentiates itself by using AI to conduct "discovery-based science." Zymergen uses machine learning techniques to predict promising experiments, absent an explicit ingoing hypothesis. Zymergen then uses robotics and wet lab automation to conduct these experiments automatically with minimal human

involvement. One of the biggest improvements of Zymergen's recommendation model, compared to human-recommended experiments, is its superior memory which removes recency bias or the tendency to think that recently observed trends will continue in the future, a common pitfall of human experiment design.

Through its approach of combining automation in R&D labs and machine learning to improve experiment design, the company reports significant efficiency gains: according to their internal analysis, Zymergen's experiment throughput can be roughly 10 times higher than that of a conventional lab, and with significantly shorter project timespans. For its customers, Zymergen promises to deliver improved economics through increased production rates or yield gains on existing products.

Zymergen's use of automation and machine learning is an instance in which AI is shaping the nature of work for a highly educated workforce. For instance, both manual and cognitive work previously done by highly educated scientists and researchers has been offloaded to automated systems and machine learning models. This has led to smaller project team sizes than typical conventional R&D labs. At the same time, Zymergen has increased need for data scientists and automation engineers to support these AI-related services. This case study explores the costs, risks, and benefits of Zymergen's approach of using automation and machine learning.

<sup>1</sup> A wet lab is a scientific laboratory designed to handle chemicals and avoid contamination, often built with specific equipment and requirements to reduce human contact with the chemicals. At Zymergen, the equipment and tools used in the automated wet lab include but are not limited to: liquid handling systems, robotic colony pickers, barcoders, acoustic dispensers, automated plate readers, robotic rule-based scripts, and systems or software used to operate this equipment.

<sup>2</sup> An 'AI-native' refers to a company that was founded with a stated mission of leveraging artificial intelligence or machine learning as a key enabling technology. 'AI-natives' can build infrastructure from the ground-up without the need to shift from legacy systems (e.g., on-premise to cloud-based storage).

<sup>3</sup> During the time of writing the case study in fall 2018, the company had raised \$174M. On December 13, 2018, the company announced a \$400M Series C round from multiple investors. See coverage of the announcement on [Bloomberg](#) and the [Wall Street Journal](#).

<sup>4</sup> "Fermentation Products Market by Type - Global Opportunity Analysis and Industry Forecast," Allied Market Research, June 2017.

## Key observations and findings:

- **Implementation challenges:** Even as an AI start-up, Zymergen found it necessary to focus on and manage the integration of AI tools into traditional workflows. In particular, their experiences showed that the explainability of models was closely linked to adoption. For instance, if a scientist or customer did not understand why a model recommended a certain experiment, it proved harder to obtain buy-in and adoption. The dynamics of change management for researchers and scientists involves a sense of pride in their work - understandable given that many have studied their domain for years and may resent seeing their expertise reduced to 'statistical analysis'. Moreover, sophisticated AI methodologies did not always produce better results, either because of the complications of working with "black-box<sup>5</sup> models or the limitations of extant data. Notably, productivity benefits reported by Zymergen were associated with significant upfront investments and operating costs. Automation involving high-throughput screening is expensive and Zymergen reports that a significant part of the \$574.1 million<sup>6</sup> funding raised as of 2018 has been directed towards building out the infrastructure for the company, such as the robotics and automated wet lab machinery. Additionally, Zymergen has higher consumable costs than most R&D labs, which is directly associated with running more experiments as part of its increased data generation strategy.
  - **Business and productivity impact:** Zymergen reports that its use of machine learning and automation has resulted in significant efficiency gains, largely in the form of higher throughput. According to internal analysis, Zymergen can achieve productivity gains via a tenfold increase in experiment output per week compared with conventional R&D labs.<sup>7</sup> Zymergen reports that its customers, typically operating in highly competitive, low-margin bio-manufacturing businesses, can realize economic benefits in the form of yield improvements and productivity gains.
- For on-market products (existing products that a company is improving upon), Zymergen reports that its program could see yield improvements of two to three times what in-house R&D programs can provide, reducing production costs through savings on inputs (such as raw materials) needed to otherwise achieve this yield. In addition, Zymergen reports that its programs typically deliver improvements in shorter durations than an in-house R&D strain improvement program: from roughly three to five years for a strain improvement program at Zymergen, compared with eight to ten years for its clients' programs.
- **Workforce and labor impact:** Automation and machine learning at Zymergen reflect a different approach to science and experimentation, compared with conventional R&D labs. Accordingly, at Zymergen, the nature of work, the skills required, and the composition and allocation of resources differs from those of conventional labs. Labor trends at Zymergen and its clients may be signals of broader workforce evolutions in the long term, especially with regard to biology R&D teams, biotech companies, and associated customers (e.g., agricultural commodity manufacturers, consumer electronics manufacturers). Workforce trends in the Zymergen case appear in two main groups: internally within Zymergen, and externally with its customers (both upstream R&D labor and downstream manufacturing or supply chain labor). Internally, Zymergen appears to have about 50 percent smaller core R&D teams per project than the average conventional R&D lab. At the same time, these core teams are supported by a higher number of machine learning and automation-related teams through a shared services model. Zymergen has a highly educated workforce, with about one-third of employees holding a Ph.D. Still, jobs at Zymergen require different skills compared to traditional R&D roles: there is less demand for scientists to do hands-on lab work; research associates are expected to have scientific

<sup>5</sup> In general, a black box system references a system in which only the inputs and outputs are visible; what goes on "inside" the black box is unknown or not easily explained -- the causes for this opacity vary, and may be due to technical or proprietary characteristics of the algorithm. When the inner-workings of a system remain unknown, this can raise issues of trust, transparency, fairness, and accountability.

<sup>6</sup> <https://www.crunchbase.com/organization/zymergen#section-funding-rounds>

<sup>7</sup> Output measured here as unique phenotypes generated in a wet lab per week.

backgrounds as well as data science skills, reflecting a shift away from roles in executing experiments manually. In addition, Zymergen roles typically include a requirement of specialization in multiple domains (e.g., biology and automation engineering, data engineering, or data science).

In addition, shorter total project durations have resulted in faster project staffing cycles. Zymergen stated that its highly educated workforce will be involved in stages of product development for some time to come, and it was interested in the opportunity to create more value by concentrating on tasks where humans have a comparative advantage, which could lead to more human involvement in early product development and post-development review and analysis.

Externally, Zymergen could have a potential upstream impact of replacing its customer's internal R&D teams or causing those teams to shift to a different type of work (e.g., implementation, new product development). Zymergen's reported productivity gains in fermentation from strain improvement programs could also have a downstream impact on manufacturing labor through increased labor and resource productivity or potentially forgone hiring related to a reduced need to build new manufacturing plants.

## Conclusion and lessons learned

Zymergen's business model is designed around the use of AI, machine learning, and automation to improve productivity compared to traditional R&D firms, and its use of automation and machine learning presents an interesting case through which to examine the changing nature of work for a highly educated workforce (e.g., Ph.D.s in biology or research associates). While the company was founded on the principle that both complex and mundane tasks could be automated out of the R&D workflow today, for the foreseeable future, scientists remain integral to Zymergen's core business. Nonetheless, the potential labor implications are substantial, both within the firm and throughout the firm's client and business ecosystems.

**Mixed labor implications — expanding beyond the boundaries of the firm:** At Zymergen, it is likely that in the short- to medium- term, the demand for scientists will continue, although the nature of their work may require more interdisciplinary technical skills. In addition, although the number of scientists required per project declines, Zymergen expects the overall demand for highly-trained support staff (e.g., data scientists and data engineers) with quantitative skill sets to continue to increase. However, demand for research associates or lab technicians may be lower in the long run. In addition to workforce changes within Zymergen, the case suggests that there could be external workforce ripple effects. For instance, Zymergen might impact the size or role of customer's internal R&D teams. In this sense, analyzing direct labor-force implications within one organization may be insufficient; as more companies and industries begin to increase their productivity through AI, there may be workforce ramifications that were not present at individual sites, and thus lie outside the traditional scope of labor impact investigations at the individual firm level.

## Open questions and future research

The Zymergen case poses a number of questions: With a large portion of manual work conventionally done by research associates and scientists offloaded to automated systems, will the more cognitively-driven tasks also be replaced by AI systems or will the new technologies be complementary? Similarly, as automation enters into the most manual and repetitive human tasks (like pipetting), what will be the trend for less repetitive, more novel tasks? How will human agility for learning compare with the time needed to design and maintain automation systems for varied tasks? Finally, what will be the cascading impacts of AI on *external* companies and industries? Given Zymergen's relatively small employee count, these may be the most profound implications of its reported productivity enhancements and merits further research. While many impacts and ramifications of AI-related technologies on labor and the economy remain unknown, the case of Zymergen calls attention to trends that suggest the need for further study.

# Appendix

## Definitions and terms used

While we acknowledge that there is no consensus on the definition of terms such as AI and automation, we would like to explain how these terms are used in the compendium:

**Artificial intelligence/AI** is a notoriously nebulous term. Following the [Stanford 100 Year Study on Artificial Intelligence](#), we embrace a broad and evolving definition of AI. As Nils J. Nilsson has articulated, artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment. (Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, UK: Cambridge University Press, 2010).

Our definition of **automation** is based on the classic human factors engineering definition put forward by Parasuraman, Sheridan, and Wickens in 2000: <https://ieeexplore.ieee.org/document/844354>, in which automation refers to the full or partial replacement of a function previously carried out by a human

operator.<sup>8</sup> Following Parasuraman et al.'s definition, levels of automation also exist on a spectrum, ranging from simple automation requiring manual input to a high level of automation requiring little to no human intervention in the context of a defined activity.

**Explainable AI** or **Explainability** is an emerging area of interest in communities ranging from DARPA to criminal justice advocates. Broadly, the terms refer to a system that has not been “black-boxed,” but rather produces outputs that are interpretable, legible, transparent, or otherwise explainable to some set of stakeholders.

In this compendium, a **model** refers to a simplified representation of formalized relations between economic, engineering, manufacturing, social, or other types of situations and natural phenomena, simulated with the help of a computer system.

<sup>8</sup> Our definition draws on the classic articulation of automation described by Parasuraman, Sheridan, and Wickens (2000): <https://ieeexplore.ieee.org/document/844354>



**PARTNERSHIP ON AI**