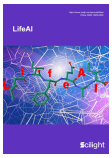
LifeAI
https://www.sciltp.com/journals/lifeai



Article

3P Framework (Prompt, Persist, Produce): Accelerating Skill Development and End-to-End Results through Human-AI Collaboration

Neal Cody [1,†], Veda Sripada [1,2,†], Katherine Papciak [1], Isaac Oyediran [1], Amanda F. Yanke [1], Bakhtawer B. Baloch [3], Chien-Hung Lu [1], Jameson Born [1], Nwe N. Aung [1], Prakruthi Harish [1], Rishav E. S. Dasgupta [1], Sashank B. Narayan [4], Gurkirat Sekhon [5], Natali Kolker [6], Nisha Singh [3], Oleksandr Romanenko [6] and Eugene Kolker [4,6,7,8,9,*]

[1] Grossman School of Medicine, New York University, New York, NY 10016, USA
[2] Chan School of Public Health, Harvard University, Boston, MA 02115, USA
[3] School of Professional Studies, New York University, New York, NY 10003, USA
[4] Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA
[5] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
[6] 1Ekaroni, Fair Lawn, NJ 07410, USA
[7] Expedaite, Jersey City, NJ 07310, USA
[8] Sync.MD/Vyrty, Anderson, SC 29622, USA
[9] School of Global Public Health, New York University, New York, NY 10003, USA
* Correspondence: eugene@kolker.ai
† These authors contributed equally to this work.

How To Cite: Cody, N.; Sripada, V.; Papciak, K.; et al. 3P Framework (Prompt, Persist, Produce): Accelerating Skill Development and End-to-End Results through Human-AI Collaboration. *LifeAI* **2025**, *1*(1), 3.

Received: 29 May 2025
Revised: 30 June 2025
Accepted: 5 November 2025
Published: 17 November 2025

Abstract: This work demonstrates how Human-AI collaboration can substantially accelerate advanced skill development and end-to-end results (transition "*from learning to doing*") in complex fields that traditionally require extensive, lengthy training. In a graduate-level course "*AI, Generative AI, and Data Science for Biomedical Informatics*" taught by Eugene Kolker, the experimental group with little to no prior experience rapidly progressed from instruction to producing credible, real-world results. Enabled by Generative AI (GenAI), the experimental group members processed data, generated code, interpreted results, synthesized literature, and documented findings over a focused 20–hour implementation phase. Applying these sophisticated skills to a rigorous, multi-step analysis of eukaryotic enzymes, they observed underlying preferred sizes, consistent with and extending upon prior work. Notably, the more experienced control group failed to reproduce these results even after 50% more time. This striking success-failure differential was enabled by the **3P (Prompt, Persist, Produce)** framework. 3P is a straightforward, robust, and reproducible Human empowerment methodology that draws its inspiration from the Socratic Method, Experiential Learning Theory, and Systems Thinking. The 3P approach is based on three pillars: (1) optimally structured communication, (2) Human expert facilitation ("player-coach" function), and (3) GenAI systems. Comparison of the results obtained by the experimental group *with* 3P (and GenAI) *versus* control group *without* 3P (and GenAI) revealed that 3P significantly shortens the time from question to insight and from learning to outcomes. Through Human-AI collaboration, 3P accelerates advanced skill development and end-to-end results, including, for instance, the preparation of this manuscript. The 3P "*from learning to doing*" paradigm not only transforms traditional training curves, but also offers a rapid, scalable, field-agnostic blueprint for accelerated research and development, innovation, workforce upskilling and reskilling, and measurable, end-to-end results.

Keywords: Human–AI collaboration; generative AI (GenAI); large language models (LLMs); small language models (SLMs) and frontier models; prompt engineering and structured prompting; 3P framework (Prompt, Persist, Produce); AI explainability; reproducibility and validation (ChatGPT, Claude, Perplexity, Cursor); AI-accelerated skill development; workforce upskilling; R&D and end-to-end results

1. Introduction

1.1. From the Experiential Learning, Socratic Method, and Systems Thinking to GenAI and 3P

The 3P framework draws its inspiration from three historically enduring approaches. First, the Experiential Learning Theory states that skills are best mastered by doing, not merely by hearing or seeing [1–5]. This learn-by-doing approach has very deep historical roots, most notably traced to the Confucian philosopher Xunzi (Xun Kuang), who lived in the third century B.C.E. A simplified version of his teaching captures the essence: "*When I hear it, I forget it. When I see it, I remember it. When I do it, I know it*" [6]. Second, the classic and modern Socratic Method complements this by structuring the "doing" around iterative questioning and dialogue, ensuring clarity, reflection, and refinement, and serving as the cornerstone of the modern conversational, structured communication [7–12]. Third, closely connected with Systems (Integrative) Biology [13–19], Systems Thinking situates both action and dialogue within a holistic method, connecting immediate tasks to broader goals and reproducibility [20–23]. Together, these three inspirations served a baseline for creating a single, synergistic 3P (Prompt, Persist, Produce) framework designed to supercharge experiential learning, skill development, and delivery of results (Figure 1).

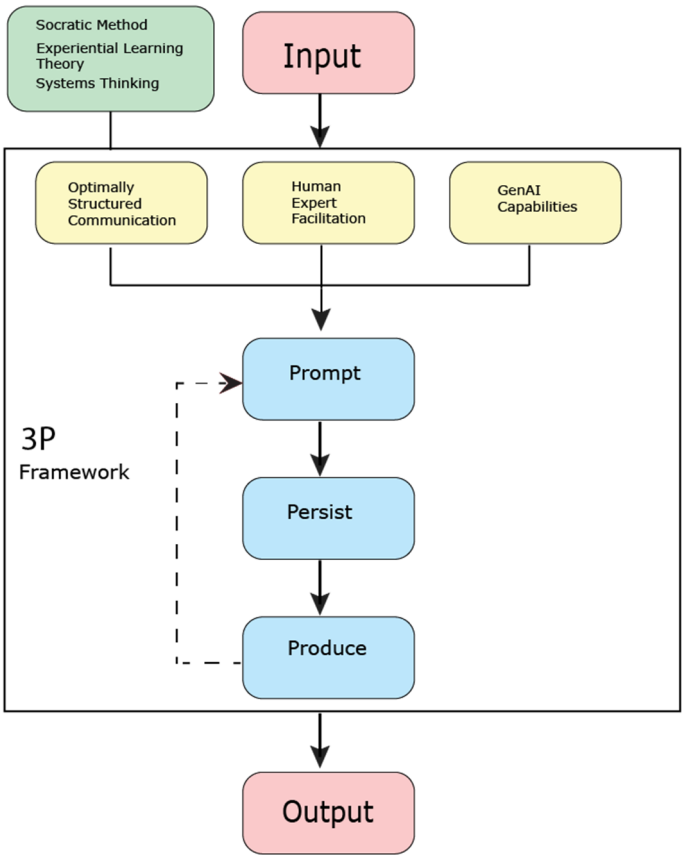


Figure 1. Overview of the 3P (Prompt, Persist, Produce) Framework.

As AI marked its 70th anniversary in August 2025 [24], advanced AI tools enable unprecedented opportunities to accelerate and transform both learning and execution. The integration of Generative Artificial (or Augmented, Applied, Added) Intelligence (GenAI) [25–28] into learning is redefining the pace, scale, and accessibility of traditional training. Since ChatGPT launch on November 30, 2022 [29], GenAI, Frontier models, Small language models (SLMs), and Large language models (LLMs) [30–34] like GPT-4o Omni [35] and Claude

3.7 Sonnet [36], and advanced AI tools like Perplexity [37] have ushered in a new era of AI-assisted research and development [38]. These modern models now go beyond text generation to assist with data analysis, hypothesis testing, and knowledge synthesis. Among others, they can empower researchers and novices alike to iterate across different fields and enable flexible, hypothesis- and data-driven developments [39–50]. More broadly, GenAI systems have brought substantial advancements to multiple, diverse Human activities, including, for instance, augmentation and automation of routine tasks, code generation, education, and advanced searches and summarization [51–65].

Specifically in biomedical research and development, AI has already demonstrated remarkable capabilities. Examples include protein structure prediction (e.g., RosettaFold, AlphaFold) and functional annotation (e.g., GrAPFI) [66–71]. Notably, the 2024 Nobel Prize in Chemistry was awarded half to David Baker for computational protein design [66,67], and the other half to Demis Hassabis and John Jumper for their development of AlphaFold's protein structure prediction [68,69]. Yet, the AI potential for guiding open-ended, exploratory pattern discovery, especially in complex fields with subtle, underlying signals requiring rigorous, multi-step analyses, remains underexplored. In these contexts, emphasizing Human-AI collaboration, including potential synergy between structured communication, Human insights and intuition, and AI-driven pattern recognition and delivery, may be particularly powerful.

In this work, we developed and tested a novel 3P (Prompt, Persist, Produce) framework. 3P is a straightforward, robust, and collaborative Human empowerment framework designed to significantly shorten the time to skill development and end-to-end results. The 3P methodology is based on three pillars: (1) optimally structured communication, (2) Human expert supervision, and (3) GenAI capabilities (Figure 1). It is worth noting that optimally structured communication is defined by the six key characteristics: it has to be clear, specific, concise, structured (e.g., STAR- or SCR-like), conversational, and optimal (Table 1). As such, it is aimed at both Human-to-Human and Human-to communication. Through its real-world results-driven approach, 3P employed structured prompting with GenAI to accelerate training and execution of the experimental group. The synergy between effective communication, Human expert facilitation, and GenAI systems enabled the experimental group members with little to no skills not only to learn, but also to apply newly obtained skills in a complex, multi-stage project. As a result, 3P significantly accelerated both the skill development and end-to-end results.

Table 1. Description of the optimally structured communication [1].

No.	Key Element	Explanation
1	Clear	One-sentence BLUF [2] takeaway easy to understand and remember
2	Specific	Names, numbers, dates, actions, and constraints
3	Concise	Remove nonessential words, targeting ≤4 points (100 words)
4	Structured	Consistent layouts: From takeaway and facts to actions and timeline
5	Conversational	Plain language, personalized active voice, and minimal jargon
6	Optimal	STAR- or SCR [3]-like, tuned for real-world outcomes with minimal back-and-forth

[1] Human-to-Human and Human-to-AI. [2] BLUF: Bottom line up front, the practice of beginning a message with its key information (the "bottom line"). [3] STAR: Situation-Task-Action-Result method; SCR: McKinsey's Situation-Complication-Resolution framework.

Within a very short 20–h implementation timeframe the experimental group completed a sophisticated, multi-step bioinformatics project from start to finish, including the detection and validation of hidden, underlying signals in eukaryotic enzyme lengths and submission of this very article. These two are the real-world, end-to-end results achieved by the experimental group, while the more experienced control group failed to do so. With 3P, both the training and execution became more inquiry-driven, dynamic, and empowering, while ensuring diligence, accuracy, and reproducibility. Ultimately, 3P aims to transform traditional learning curves and provide a scalable, field-agnostic approach for accelerated research, innovation, and workforce upskilling and reskilling.

1.2. Bioinformatics Use Case: Analysis of Eukaryotic Enzyme Lengths

There were three principal considerations in choosing the analysis of eukaryotic enzyme lengths as our bioinformatics use case for this work. First (timing and context): To produce robust, reproducible outcomes, this work utilized three advanced GenAI tools: GPT-4o Omni [35], which was released on 13 May 2024, Claude 3.7 Sonnet [36], released on 24 February 2025 (right before the project's start), and Perplexity [37]. It was carried out within the two final months (from mid-March to mid-May 2025) of the Spring 2025 semester's graduate-level course "*AI, Generative AI, and Data Science for Biomedical Informatics"* taught by Eugene Kolker at New York University. A bioinformatics analysis was both directly relevant and educationally appropriate. Second (use case): The chosen use case needed to be

complex, that is to require a rigorous, multi-step, and sophisticated analysis. Such a use case would challenge the experimental group *with* 3P (and GenAI) and control group *without* 3P (and GenAI) to implement a variety of computational and statistical methods and provide the meaningful interpretation of the obtained findings. These methods include exploratory data analysis, distribution fitting, signal processing, spectral analysis, and statistical testing [39–46], as well as biological interpretation of their observations. Third (defensibility): For the 3P framework to be robustly validated, the analysis had to involve complex (hidden, underlying) phenomena with well-established, peer-reviewed observations. The eukaryotic enzyme length distribution, including the presence of preferred sizes and their multiples (subtle, underlying signals), represent such a phenomenon: it is well supported in prior peer-reviewed literature and therefore suitable for benchmarking newly observed findings. Taken together, analysis of eukaryotic enzyme lengths addresses all three required criteria: it is biologically meaningful, computationally and methodologically demanding, and amenable to cross-validation against prior peer-reviewed work.

Various biological factors influence eukaryotic protein lengths, such as folding dynamics, evolutionary pressures, energetic effects, and gene regulation [72–77]. While a comprehensive review lies beyond this section, several meaningful underlying drivers merit mentioning. One intriguing observation in bioinformatics is that the biological sequences contain diverse "hidden codes" in addition to the canonical triplet genetic code [78,79]. Among these, the underlying "code" [79] indicates preferred lengths for protein segments (domains, modules, or subunits), typically ranging from 50 to 250 amino acids (aa) [80–82]. Specifically, preferred sizes between 100 and 150 aa (so-called "segmentation code") have been reported most consistently across multiple studies, e.g., [72,74,76–79,83–101]. Historically, the concept of typical protein sizes dates back almost a century, originally proposed by Theodor Svedberg [102,103].

Another intriguing observation in bioinformatics is that functionally important proteins are more evolutionarily conserved [104–106]. Additionally, proteins of greater length often possess multi-domain structures resulting from gene recombination and fusion (protein "domain shuffling") [80–82,107,108]. Interestingly, eukaryotic proteins display a notably high proportion of such modular architecture with at least two thirds classified as multi-domain [77,109–113]. Then, methionines, as start residues, would potentially mark those locations in multi-domain proteins that correspond to the potential DNA recombination sites, the borders between the fused genes [114]. Indeed, the analysis of eukaryotic protein sequences showed that methionine residues preferentially appear at the positions corresponding to the multiples of the preferred size of ~125 aa [88,114]. In total, several important factors combined together to produce the observed protein length distributions, including, for instance, structural (protein folding constraints), sequence ("segmentation code" of protein subunits or domains), evolutionary (functionally important proteins are more evolutionary conserved; gene recombination and fusion driving "domain shuffling"), and energetic (multi-domain proteins prevalence in eukaryotes) effects [72–101,104–114].

All these, in turn, make evolutionary conserved eukaryotic proteins a meaningful target. One class of such diverse, evolutionary conservative proteins is eukaryotic enzymes, making them an attractive use case [72,74,87,88]. Clearly, most eukaryotic enzyme lengths follow the background distribution. Still, we model the overall length distribution as a statistical mixture of this background distribution combined with n individual peaks at the potential preferred (subunit) length μ and its multiples (at $2 \times \mu$, $3 \times \mu$, … $n \times \mu$). Additionally, analysis of eukaryotic enzyme lengths with this potential hidden, underlying signal represents a great testbed for the 3P framework because it provides: (1) freely available, diverse, large enough datasets, (2) example of the complex, multi-stage project to detect hidden, underlying signals, (3) need for different sophisticated statistical and signal processing methods rarely accessible to novices, (4) clear, robust validation pathways via comparison to established peer-reviewed literature, and (5) opportunities for meaningful biological interpretation of computational results. Therefore, this work hypothesized that 3P can enable rapid, rigorous, and validated investigation and interpretation of a computationally and biologically challenging analysis of eukaryotic enzyme lengths within a very short timeline.

2. Methods

2.1. 3P Framework versus Traditional Approaches

Let us introduce here the 3P (Prompt, Persist, Produce) framework. 3P drew its inspiration from three distinct sources: the classic and modern Socratic Method (including critical thinking, Neo-Socratic Method, active learning and listening skills, and reasoning, see [1–5]), Experiential Learning Theory, and Systems Thinking (Figure 1). 3P is designed to accelerate skill development and end-to-end results by combining the power of GenAI systems with iterative, effective communication and Human expert actionable guidance. Prior to the 20–hour focused implementation timeframe, the Human-to-Human optimally structured communication was instrumental to describe the project's expectations, goals, and deliverables (Table 1, "Input" in Figure 1). This communication also concisely

introduced prior peer-reviewed literature and key approaches to be implemented, focusing on the required reading of only two articles [72,74] (other relevant publications were mentioned, but were not required to read).

First (Prompt, optimally structured conversation): From the project's start the experimental group (described in detail in "Experimental *versus* Control Group" section below) also employed the optimally structured Human-to-GenAI communication (Table 1). Specifically, the experimental group started to develop a set of complementary structured prompting approaches. They include: single-paragraph summaries, detailed data processing instructions, a comprehensive analysis pipeline, targeted statistical method prompts, and technical spectral analysis prompts. Second (Persist, iterative improvements under Human expert facilitation): Each specific prompt type was initially submitted to both GPT-4o Omni and Claude 3.7 Sonnet [35,36] by the experimental group consisting of three teams. The experimental group then improvised and iterated with three GenAI tools (GPT-4o, Claude 3.7, and Perplexity [37]) until more optimized, reliable, and reproducible results were achieved. Third (Produce, end-to-end results): Three teams' cross-comparison, competition, and collaboration resulted in a much faster iterative process towards significant overall improvements and delivering the real-world outcomes (Feedback Loop in Figure 1, Table 2). At this stage, all three different GenAI systems were used both for literature search, summarization, and manuscript preparation ("Output" in Figure 1). As a result, GenAI capabilities became the P3's integral pillar that is also utilized in every Produce-phase activity.

Table 2. Experimental group (with 3P): Team composition, methods, and results.

Category	Team A	Team B	Team C
Background Composition *	3 Biology, 1 Computer Science	1 Biology, 1 Business, 1 Engineering	4 Biology
Programming Language	Python (NumPy, SciPy, Pandas)	R (*tidyverse*, spectral package)	Python + R
AI Prompting Approach	Detailed step-by-step instructions with specific statistical parameters	Open-ended queries with iterative refinement focused on methodology	Domain-specific prompting with biological context
Dataset Preprocessing	Extensive taxonomic grouping, duplicate removal (reduced from 18,076 to 15,103 unique enzymes)	Creation of a non-redundant dataset by selecting max length representative per unique enzyme	Different filtering strategies filtering out redundant protein names (Rule 1) and redundant sequences (Rule 2)
Primary Method	cos-Fourier Transform with signal smoothing	Spectral Analysis of Distribution (SAD) method, per OMICS 2002	Extended SAD with multiple filtering rules
Statistical Methods	Chi-square test, Gaussian Mixture Model (GMM), LOWESS, Wavelet analysis	Likelihood ratio test (LRT), SAD, Normal mixture model	LRT, GMM + Normal mixture model, AIC/BIC model comparison
Results	Eukaryotic enzymes have preferred sizes of ~117–127 aa and their multiples that are non-random due to diverse effects	Eukaryotic enzymes of ~124–137 aa typical lengths and their multiples are evolutionarily preferred	Eukaryotic enzymes exhibit preferred sizes of ~121–131 aa and their multiples due to structural constraints

* The background composition of the control group: 4 Computer Science, including 1 with Bioinformatics experience.

Two other integral pillars of the 3P methodology (Figure 1) are the optimally structured communication and sustained, hands-on guidance from a Human expert facilitator, functioning in parallel as advisor, mentor, and player-coach. In the 3P implementation, effective Human-to-Human communication served as an intrinsic, vital core. Similarly, the importance of such effective Human-to-GenAI communication is hard to overstate. Optimally structured communication, as defined in Table 1, represents a reasonable "optimal" compromise tailored to both Human-to-Human and Human-to-AI collaboration. In addition, at every 3P phase, the Human expert (here the professor) provided real-time feedback, targeted corrections, and strategic direction. This supervision encompassed conceptual guidance, methods correction, critical appraisal of intermediate results, and best-practice reinforcement. Far from being peripheral, this "player-coach" function proved essential for accelerating three critical dimensions: 1. learning, 2. the transition from learning to practical execution ("from learning to doing"), and 3. the execution itself. Notably, such active supervision is not merely a feature of this work, but a crucial ingredient for successful 3P implementations. The 3P approach's effectiveness is fundamentally anchored in the above iterative, expert-driven mentorship, which seamlessly integrates Human insights and GenAI capabilities.

2.2. 3P Hypothesis

In this work, we hypothesize that the 3P framework, anchored in the effective communication and GenAI prompting, iterative approach, and Human expert facilitation, can dramatically accelerate the training and execution in complex fields. Specifically, we expect that within the constraint of the 20–hour implementation window, the experimental group could: (1) learn, comprehend, and apply diverse bioinformatics, statistics, spectral analysis, signal processing, and Fourier transform methods (that traditionally demand significant expertise and prolonged study and training), (2) compare and validate their results with previous peer-reviewed biological observations, and 3. document their findings for this article submission. To evaluate this three-part hypothesis, we compared the results obtained by the experimental group using 3P (and GenAI) *versus* the control group of more experienced peers restricted from 3P (and GenAI). This design enabled the direct test of whether 3P-empowered participants could match or even surpass the outcomes of traditionally trained and more experienced peers within the same short implementation timeframe. The classic A/B comparison between the experimental and control group results combined with the testing and validation of the above three-part hypothesis would indicate whether the 3P framework (Figure 1) succeeds in accelerating skill acquisition and producing robust, reproducible, real-world, and end-to-end results.

Additionally, all the participants of the experimental group evaluated pre- *versus* post-project confidence assessments based on the five following complementary criteria. First, Completeness measured coverage of all essential methodological components on a 1-5 scale. Second, Technical Accuracy assessed the correctness of described techniques and parameters, also on the same scale. Third, Clarity evaluated logical flow and accessibility to the target audience on the same scale as well. Outputs scoring from 1-5 on Completeness, Accuracy, and Clarity were considered publication-ready with minimal human editing. Fourth, Iteration Efficiency tracked the number of prompt refinements needed to achieve the desired output, with lower numbers being better. Fifth, Novel Insight Generation captured unexpected but valuable methodological contributions.

2.3. Design and Composition: Experimental versus Control Group

To address validity and reproducibility of the proposed 3P framework, we randomly divided the experimental group of 11 students into three teams: Teams A and C with four members, and Team B with three members. This size ensured adequate distribution of backgrounds per team (Table 2) and allowed meaningful cross-validation across three complementary pipelines. These team assignments reflected the following backgrounds: 8 in biology, 1 in business, 1 in engineering, and 1 in computer science, with none of bioinformatics training (see Table 2 for the teams' specific compositions). The entire project, with optimally structured communication-, Human supervision-, and GenAI systems-empowered 3P, was performed in parallel by these three teams over the *20–hour implementation phase.* The teams were not only able to cross-validate their findings, but also, through close and timely collaboration and healthy competition, shared insights and supported one another. In addition, strategies proven effective by one team were adopted by others, creating synergy and amplifying both the project's learning and implementation outcomes. This process significantly improved overall results and accelerated the learning, the transition from learning to practical execution, and the execution experience for the participants of the experimental group.

The absence of an initial specialized control group, using traditional approaches without 3P, posed a severe limitation for isolating the 3P contributions. To overcome this, we introduced the control group, instructing them to conduct their own 20–hour implementation phase *without* 3P. Purposely by design, the control group was more experienced and better equipped for the required multi-step, complex project with four computer scientists, and one of them having bioinformatics experience. To reiterate, the experimental group members came primarily from a non-computational background (Table 2).

Representative prompts from Teams A, B, and C are presented in Tables S1 and S2, also accessible via the GitHub (https://github.com/nyu-vilcek-bmi/BMI_GenAI_Class_2025_LifeAI_Paper, accessed on 14 November 2025) alongside the corresponding scripts and summary outputs generated by each team. All three teams meaningfully contributed at every stage of the project. Notably, Team C's overall methodology was the most comprehensive and most closely aligned with the project's research phases (see Table S1). For consistency, the work produced by Team C was used for illustrative purposes throughout the entire manuscript (Tables 2 and S1, Results section). For completeness, sample prompts from Teams A and B are also included in Table S2.

2.4. Parallel Analysis by Experimental Group

This project employed a three-team parallel analysis design for the experimental team focusing on the bioinformatics use case, described in the Section 1. Let us also briefly reiterate key assumptions and uncertainties of the multi-step analyses implemented in this work. First, the original data might reflect some biological

mechanisms that could generate some periodic components. Second, along with the underlying periodic components, the observed data include some background caused by a different origin, without any major systematic biases. Third, the total number of data points is large enough to provide a representative sampling of the random uncertainties. Fourth, per the sampling theorem [115], more than two data points must exist per period of the resolvable oscillating component. Here, each team independently implemented complementary methodological strategies to investigate potential preferred sizes in eukaryotic enzyme length distributions. All three teams were analyzing the same datasets in parallel, providing methodological diversity and enabling robust cross-validation.

Diverse eukaryotic proteins were retrieved from the UniProt database [116] via the UniProt REST API and Python scripting (a revised Dataset 1, Table 2). API calls removed fragments and incomplete proteins while retaining entries with enzyme classification numbers (EC numbers). Data pre-processing procedures included column standardization, numeric conversion of protein lengths, and restricting analysis to proteins between 50–600 aa (with very short and long entries removed). Additionally, mitochondrial proteins were removed, while entries with demonstrated evidence at the protein level were retained in a new Dataset 2 (Figure S1A, Table 3). Then, further filtering steps [72,74] were applied to both initial Datasets 1 and 2 resulting in their significant reduction. The processed Datasets 1 and 2 were reduced to 14,685 and 13,328 eukaryotic enzyme sequences corresponding to ~55.4% and ~55.9% of their initial sets, respectively (Table 3).

Table 3. Four datasets: Number of eukaryotic enzymes.

Dataset	Initial	Processed	% *
Dataset 1	26,486	14,685	55.4
Dataset 2	23,831	13,328	55.9
Dataset 3	27,600	14,881	53.9
Dataset 4	21,813	12,378	56.7

* Percent of Initial dataset.

We developed a combined statistical approach to determine whether eukaryotic enzymes exhibit preferred lengths. This framework integrated two top lines of evidence: (1) distribution non-uniformity assessed via chi-square tests against a uniform null distribution, and (2) multi-modality evaluated through optimal component selection in mixture models [117,118]. Additionally, the significance of observed signals was assessed using likelihood ratio tests (LRT) and statistical testing of spectral peaks. Model comparison was performed using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to evaluate hidden, underlying periodic models *versus* background-only models [119–122]. To rigorously evaluate potential oscillation, we employed maximum likelihood estimation (MLE) to fit both the Gamma Mixture Model (GMM) and Normal mixture model to the eukaryotic enzyme length distributions [123–125]. The Gamma component captured the background model, while one or more Normal components (the signals under investigation) represented potential periodic peaks aligned at integer multiples of a potential preferred length μ [126]. We estimated model parameters by maximizing the log-likelihood using numerical optimization. LRT statistics were computed as twice the difference in log-likelihoods and evaluated against a chi-square distribution [118,125]. The AIC and BIC analysis guided model selection, balancing fit and complexity to avoid overfitting, while identifying genuine (biological) signals [119–122,127].

Specifically, Team A focused on multiple spectral analysis approaches including cosine-Fourier transforms, wavelet analysis, and advanced spectral methods. Team B emphasized a reproduction of the previously published Spectral Analysis of Distributions (SAD) methodology [72,74,127–137] using R-based statistical methods with LRT and GMM. Team C applied exploratory data analysis [138], similar SAD algorithms, and diverse statistical modeling (LRT, GMM, Limited-memory BFGS with Bounds, and AIC/BIC analysis). Validation strategies included cross-taxonomic testing via Kolmogorov–Smirnov test [139], cross-validation by bootstrap confidence estimation, and consensus analysis across multiple signal processing methods. The combination of these independent approaches created a comprehensive analysis pipeline validating findings across multiple statistical and computational approaches (Table 2).

2.5. Comprehensive Reanalysis

To respond to reviewer recommendations, we carried out a complete, comprehensive reanalysis that involved both the experimental and control groups. This work focused on three new datasets: a new Dataset 2 (described in the previous section) and two new sets Datasets 3 and 4. For the initial Dataset 3, we pulled 27,600 eukaryotic enzyme sequences from the UniProt database [116] using the UniProt API shown in Figure S1B. The selection

was limited to the Eukaryota taxonomy (ID: 2759) and further filtered to include only proteins with EC numbers, reviewed status, and evidence at the protein level (Table 3). Additionally, we removed duplicated entries from Dataset 3 to get 21,813 eukaryotic enzyme sequences in the initial Dataset 4 (Table 3). To significantly reduce redundancy, while keeping functional diversity, we further processed the Datasets 3 and 4 with the Clusterize function from the DECIPHER Bioconductor package [140]. Clusterize applies a three-phase heuristic algorithm [141]: sequences above the similarity threshold (different thresholds were explored, with the distance cutoff 0.15 chosen for the analysis; 0.15 corresponds to ≥85% similarity, calculated as the fraction of differing positions in pairwise alignments within overlapping regions) were grouped into existing clusters, while more divergent ones started new clusters. We then applied the same additional filtering steps used earlier for Datasets 1 and 2. As a result, the processed Datasets 3 and 4 were reduced to 14,881 and 12,378 eukaryotic enzyme sequences corresponding to ~53.9% and ~56.7% of their initial sets, respectively (Table 3). Finally, to strengthen reproducibility, we also cross-validated the obtained results with Cursor assistance [62].

3. Results

3.1. Experimental versus Control Group Outcomes

The experimental group, working with the 3P framework progressed rapidly from instruction to producing statistically and biologically validated results within the 20–hour implementation window. The experimental group was able to successfully address the entire three-part hypothesis. The more experienced control group failed to do so even with 50% more time. Specifically, all three teams of the experimental group were able to: (1) comprehend and apply diverse bioinformatics, statistical, spectral analysis, signal processing, and cosine-Fourier transform methods, (2) compare and validate the observed results with previous biological peer-reviewed literature, and (3) fully document their findings resulted in the original submission of this very article.

There is a complementary way to look at 3P and the above three-part hypothesis by focusing on three critical questions: (1) learning question: can the 3P framework (that is GenAI with optimally structured communication and Human expert facilitation) accelerate the transition from instruction to rigorous analysis and real outcomes? (2) validation question: can the experimental group with 3P and the control group without 3P produce reproducible results consistent with the peer-reviewed findings of eukaryotic enzymes' preferred sizes (~125 aa and its multiples)? and (3) communication question: can the experimental and control group document an optimally structured communication (Table 1) of both their methods and results at a publication-ready level within the above extremely short timeframe?

As a result, the experimental group positively answered all these three questions within the 20–hour implementation timeframe, while the control group was unable to do the same. "By design", the control group was better equipped for the required, sophisticated, multi-step project, having four computer scientists and one of them with bioinformatics experience. Notably, the members of the experimental group (Teams A–C) were primarily from non-computational backgrounds (Table 2). Interestingly, the more experienced control group spent 10 additional hours (total of 30 hours), that is about 50% more time but still *failed* to produce outcomes comparable in scope and quality to those of the experimental group. Subsequently, we *estimated* that the control group would have to spend in total between 20–to–40 *additional hours* to succeed. In other words, the control group would require 100%–to–200% more total time than the experimental group to match the same end-to-end results.

3.2. Experimental Group: Pre- versus Post-Project Confidence Assessments

We assessed the experimental group's prior experience with three statistical and signal processing methods: Spectral Analysis, Gamma Distribution, and Fourier Transform. Three categories (No Experience, Basic Understanding, or Some Practical Experience) of prior experience and understanding were assessed across these three methods by this survey. The X-axes represent the experience categories, while the Y-axes show the percentages of respondents within each (see Figure 2). Not unexpected, but still it is worth noting that the experimental group participants reported neither prior experience nor prior understanding of Spectral Analysis (Figure 2A). Approximately 27% reported a basic understanding, but no prior practice with Gamma Distribution (Figure 2B). For Fourier Transform, ~18% reported a basic understanding and another ~18% reported limited prior experience, while nearly two-thirds had none (Figure 2C). Furthermore, the pre- and post-course 1–5 scale surveys demonstrated substantial confidence gains across all concept and skill areas (Table 4). Confidence in Spectral Analysis increased by 75% (from 2.0 to 3.5), Biological Interpretation by >57% (2.1 to 3.3), Statistical Modeling by 48% (2.5 to 3.7), and GenAI utilization increased by >27% (3.3 to 4.2; Table 4). This analysis showed how the 3P framework compressed the typical learning curve for sophisticated bioinformatics work described here (Table

4). In turn, this revealed that 3P enabled quick, parallel, and simultaneous mastering of diverse sophisticated concepts and skills and their implementations to deliver end-to-end results. Overall, 3P enabled a novel, accelerated experience that is distinctively different from traditional, lengthy, and sequential training approaches.

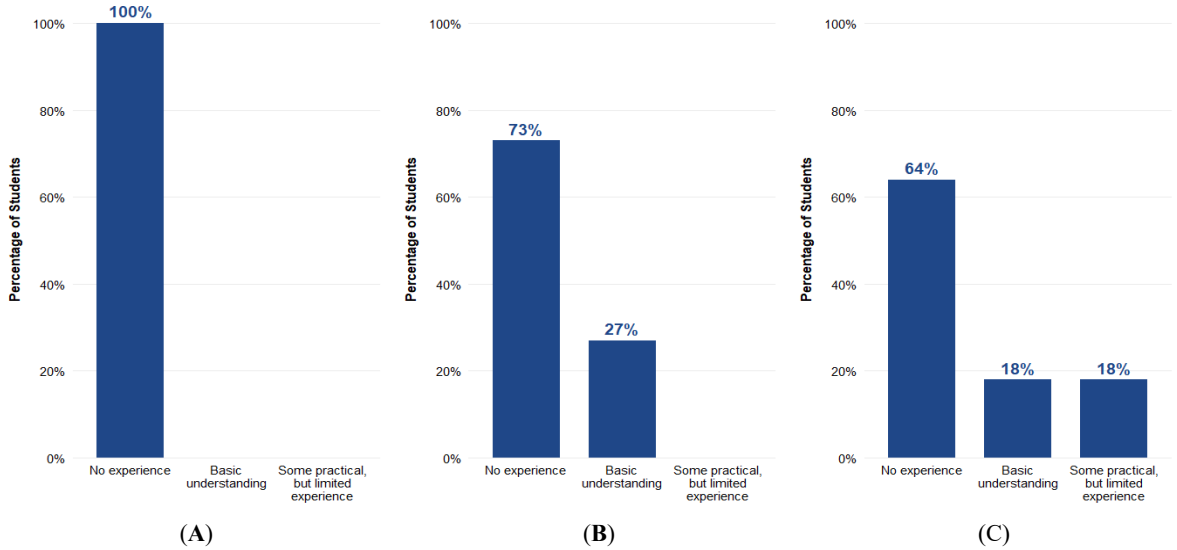


Figure 2. Experimental group: Prior experience and understanding levels across statistical and signal processing methods. (**A**): spectral analysis experience, (**B**): gamma analysis experience, (**C**): Fourier transform experience.

Table 4. Experimental group: Confidence metrics pre- *versus* post-project *.

	Pre-Course		Post-Course		
Skill Area	**Mean**	**Confidence Level**	**Mean**	**Confidence Level**	**Average Gain**
SAD Implementation	2.0	Low	3.5	Moderate-High	+1.5
Biological Interpretation	2.1	Low	3.3	Moderate	+1.2
Statistical Modeling	2.5	Low-Moderate	3.7	High	+1.2
GenAI Utilization	3.3	Moderate	4.2	High	+0.9

* Confidence levels based on scale: 1 = "Non-confident", 2 = "Low confident", 3 = "Moderate confident", 4 = "High confident", and 5 = "Extremely confident".

Furthermore, we evaluated performance of three prompt metrics (Specificity, Quality, and Iterations) across four distinct sessions of the initial submission of this article: Introduction, Methods, Results, and Discussion & Conclusions. The X-axis categorizes the data by an article section, while the Y-axis displays the corresponding scores on a standardized scale (see Figure S2). Specificity and Quality are measured on a 1-5 rating scale, with higher scores indicating better performance. Iterations represent the count of prompt refinements needed to achieve desired outputs, with lower scores indicating better efficiency. As a result, the prompt evaluation metrics indicated that high-quality, methodological, real-world outputs were generated after several iterations (Figure S2). Still, these estimations were obtained for the initial submission of this article. For the resubmission, the usage of GenAI was more targeted, in line with recommendations by the recent review [142].

3.3. Detection of Preferred Sizes of Eukaryotic Enzymes

For the analysis of the eukaryotic enzyme lengths, three complementary approaches were developed by Teams A, B, and C (Table 2). It provided a comprehensive method with Team A's prescriptive investigation and Team B's generalized SAD and cross-validation approaches. For instance, specifically for the processed Datasets 1 and 2 (Table 3), Teams A and B detected preferred sizes of ~117–129 aa and ~124–137 aa, respectively (Table 2). Again, for consistency, we illustrated our findings, based on the analysis produced by Team C. Their approach combined Python and R programming with domain-specific biological context prompting. Team C employed SAD with multiple filtering strategies, Fourier Transform, Likelihood ratio test (LRT), and complex Gamma Mixture Models (GMM) and Normal Mixture Models for cross-validation. For instance, their results for the same data revealed preferred sizes of ~121–132 aa (Table 2). Distributions of the eukaryotic enzyme lengths for the processed variants of Datasets 2 and 4 (two smallest, most cleansed datasets) are shown in Figure 3A,B, respectively. The

smoothed versions of these distributions show maxima at ~125aa multiples, corresponding to multi-domain proteins prevalent in eukaryotes [77,109–112].

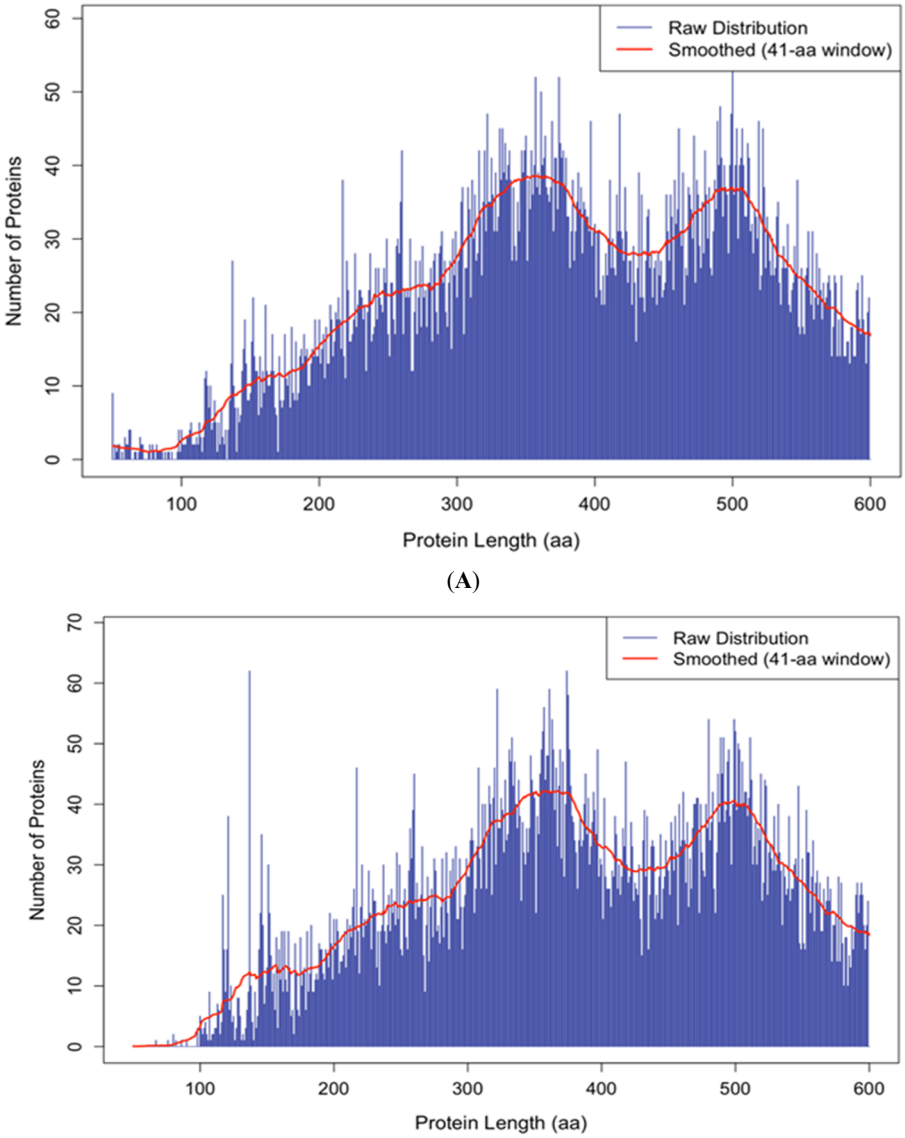


Figure 3. Distributions of eukaryotic enzyme lengths. (**A**): processed Dataset 2, (**B**): processed Dataset 4. The histograms correspond to the actual (raw) distributions, and the smoothed curves correspond to the running averages with a 41-aa window.

These maxima were also detected by the SAD approach represented by the cosine spectra in Figure 4A,B, respectively. In these cosine spectra maxima at ~125 aa showed the highest amplitudes. (Also, there are numerous other maxima with lower amplitudes in these spectra, indicating other potential underlying signals, like, for instance, "exon shuffling" [143,144]. Analysis of these and other potential signals is beyond the scope of this work). Again, the appearance of the main signal at ~125 aa and its multiples indicates potential preferred sizes of the eukaryotic enzymes (Figures 3, 4, S3 and S4). Furthermore, the AIC/BIC analysis consistently supported the periodic mixture model across all four datasets. All four Datasets 1–4 demonstrated virtually the same results, including, for example, the estimated preferred lengths (the base periodic units) μ of ~127.9 aa and ~127.7 aa, for the two most cleansed, processed Datasets 2 and 4, respectively (Table 3).

Additionally, the statistical significance of the observed main signals was utilized by the LRT, GMM, and Normal Mixture Models. The estimated parameters for all four processed Datasets 1–4 (Table 3) are summarized in Table 5. The estimated model parameters included the preferred lengths (the base periodic units) μ ranging from ~122.48 aa (Dataset 4, Table 5 and Figure 5B) to ~129.04 aa (Dataset 2, Table 5 and Figure 5A) and σ (for the normal distribution of a single unit size) ranging from ~15.7 aa (Dataset 3) to ~17.71 aa (Dataset 2). While strengths of the

individual peaks at 1×, 2×, 3×, and 4× μ are different, ranging from ~0.001 to ~0.107 (Table 5, Figures 5 and S5), total sums thereof are much more similar, ranging from ~0.147 (Dataset 1) to ~0.194 (Dataset 4, Table 5). Furthermore, LRT resulted in highly significant observations, ranging from $p = 3.05\times10^{-55}$ (Dataset 3) to $p = 3.23\times10^{-33}$ (Dataset 4, Table 5, Figure S6). These values provide robust statistical confirmation of the observed underlying signals, with the latter (largest) value corresponding to a Z statistic of ~12.01 (It is worth noting that a "classic" p value of 0.05 corresponds to a Z statistic of ~1.96). Overall, similar results were obtained for all four processed Datasets 1–4 pointing to statistically significant observations of ~125 aa preferred sizes and their multiples (Table 5, Figures 5, S5 and S6). Again, all four processed Datasets 1–4 described here (see also Table 3) could be independently analyzed.

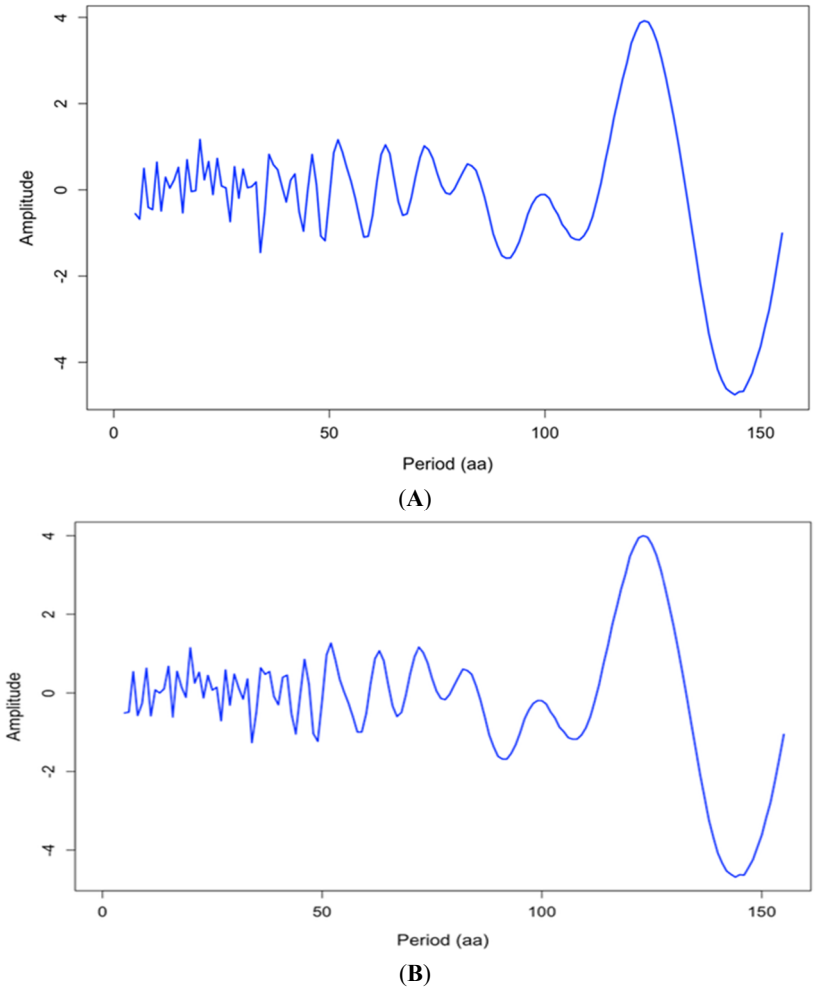
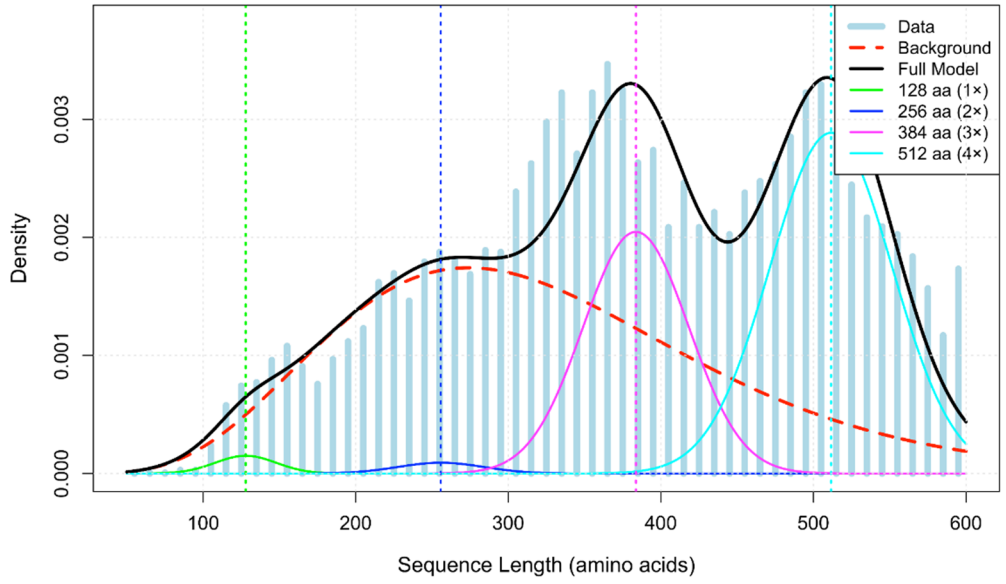


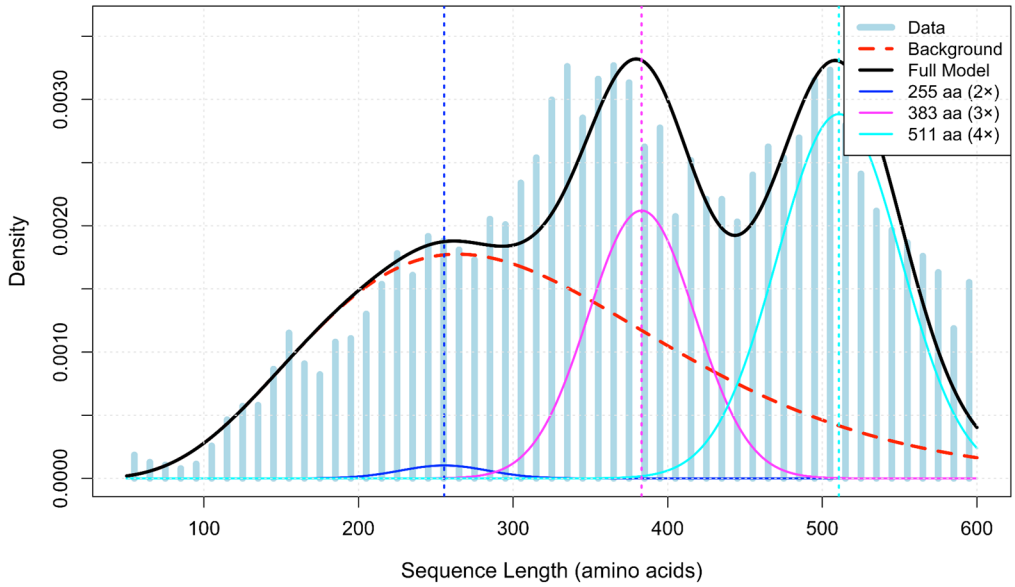
Figure 4. Spectral analysis of eukaryotic enzyme lengths. (**A**): processed Dataset 2, (**B**): processed Dataset 4. The cosine spectra generated by SAD achieve maxima amplitude at ~125 aa.

Table 5. Processed datasets: Statistical parameters and p values.

Parameter	Dataset 1	Dataset 2	Dataset 3	Dataset 4
μ	125.21	129.04	122.98	122.48
σ	15.75	17.71	15.70	17.51
p1	0.011	0.022	0.004	0.001
p2	0.001	0.001	0.006	0.017
p3	0.05	0.028	0.077	0.087
p4	0.085	0.107	0.082	0.089
sum(p1;p4)	0.147	0.158	0.169	0.194
p value	2.36×10^{-51}	1.31×10^{-50}	3.05×10^{-55}	3.23×10^{-33}



(**A**)



(**B**)

Figure 5. Statistical mixture modeling of enzyme lengths. (**A**): processed Dataset 2, (**B**): processed Dataset 4. (**A**) For Dataset 2, the mixture model combines a gamma background distribution (red dashed) and overall model (sum of normal peaks, black) with two small peaks at 1× (green, *p1*) and 2× (blue, *p2*) multiples of the base period μ and two prominent peaks at 3× (pink, *p3*) and 4× (magenta, *p4*) multiples of μ (see Table 5). (**B**) Similarly, Dataset 4 showed one small peak at 2× (blue, *p2*) and two prominent peaks at 3× (pink, *p3*) and 4× (magenta, *p4*) multiples of the base period μ (Table 5).

4. Discussion and Conclusions

4.1. From Novices to Skill Development and End-to-End Results

Under hands-on guidance by the Human expert ("player-coach"), 3P transformed the learning and doing experience that typically requires a lengthy, months-long, specialized training. The 3P "from learning to doing" experience enabled the experimental group with little to no prior background to conduct advanced research and deliver end-to-end results within the extremely short, 20–hour timeframe. This Human-AI collaboration followed the "breathing cycle": a process of continuous idea generation and refinement between the experimental group, their instructor, and AI tools. This process ensured that the AI-augmented 3P workflow remained grounded in Human interpretation and focused on the next steps towards the results delivery. The experimental group novices

quickly mastered technical methods, engaged with real-world questions and actionable insights, and produced publication-ready work, significantly lowering traditional barriers to entry for specialized complex fields.

This transformation was driven by the 3P (Prompt, Persist, Produce) framework, which promotes iterative, AI-assisted cycles of parallel advanced skill development and delivery of reproducible, robust results. Rather than progressing linearly, the experimental group teams explored ideas (Prompt), refined analyses through guided troubleshooting and (on average) two-to-four short, targeted iterative improvements (Persist), and shared results (Produce; for details, see Tables 1–5, S1 and S2, Figures 1–5 and S1–S6). Notably, the entire project, from its launch to the manuscript submission, was completed in just two months within the focused 20–hour implementation phase. Synergy between the "player-coach", optimally structured Human-to-Human and Human-to-AI communication, and GenAI contributed at all project's stages. They expanded from data processing and selecting spectral analyses and statistical tests through interpreting the observed underlying signals and validating biologically plausible mechanisms to documenting the obtained and validated results into submission of this very article. Diverse prompting strategies were evaluated to improve methodological clarity and insight generation. While GenAI enabled rapid access to methods and literature, final interpretation and validation remained a Human-driven process grounded in Systems Thinking and research reasoning [20–23,39,40,45,145].

Several factors guided the choice of eukaryotic enzyme analysis as the use case. They involved the timing, context, and defensibility, including prior, well-established, and peer-reviewed publications. Numerous important biological and evolutionary effects contributed together to produce the eukaryotic enzyme length distributions observed nowadays. Some key of these include, for example, structural (protein folding constraints), sequence ("segmentation code" of protein units or domains), evolutionary (functionally important proteins are more evolutionary conserved; gene recombination and fusion driving "domain shuffling"), and energetic (multi-domain proteins prevalence in eukaryotes) effects [72–101,104–114]. Analyses of diverse eukaryotic enzyme datasets consistently detected preferred sizes around ~125 aa, matching, among others, "segmentation code", typical independently folding domains and ribosomal pausing checkpoints (see the Sections 1 and 3). Applying SAD and complementary models to today's UniProt data [116] reproduced and extended earlier findings with statistically robust and biologically grounded updates. In addition, this project demonstrated how 3P with GenAI enables rapid learning and implementation of sophisticated spectral, signal processing, and modeling methods to achieve meaningful, validated, and robust results.

It is worth noting that the more experienced (by design!) control group was unable to match the results comparable to those produced by the experimental group within the same 20–hour timeframe. Moreover, with 10 additional hours (50% more total time), the control group still failed to obtain results of the experimental group and were estimated to require in total 20–to–40 additional hours (100%–to–200% more total time) to do so. Overall, this work undoubtedly confirmed the three-part hypothesis behind the 3P framework. In other words, 3P, anchored in optimally structured communication and GenAI prompting, iterative approach and Human expert facilitation, can indeed dramatically accelerate the training and execution in complex fields.

4.2. Current Limitations

Several specific limitations should be considered when interpreting the findings of this work. (Broader, well-documented limitations of GenAI and LLMs are beyond the scope of this section.) First, the small sample size of 16. There were 11 participants in the experimental group (Teams A–C), 4 in the control group, and 1 technical assistant (helping with the reanalysis, for details see Methods and Results). This might somewhat limit the work scalability to larger contexts. Second, conducting the work at a single institution and with a single expert (professor, instructor, "player-coach") could potentially limit generalizability to other environments. Third, the reliance on the surveys' self-reported confidence measures might not ideally reflect actual participant's competency, as these assessments are subject to a potential response bias.

Fourth, the short-term nature of the work captured immediate confidence gains but did not track long-term retention of the obtained skills. Fifth, text-based and voice-based AI approaches utilized here were also complemented by PowerPoint presentations and short videos. Still, specialized SMLs and more advanced Frontier models, deep research, multimodal, and Agentic AI approaches could be beneficial [146–160]. Despite these evolving opportunities, the current version of the 3P framework already demonstrated remarkable improvements over traditional approaches, establishing a strong foundation for its wide-spread applicability in diverse (complex) contexts, including research and development, innovation and optimization, and workforce upskilling and reskilling.

4.3. Future Directions and Broader Implications

The experimental group's compressed journey from start to publication highlights the potential of 3P and, more broadly, Human-AI collaboration, to accelerate real-world outcomes across diverse fields. Ultimately, our work demonstrates that advanced, sophisticated skills need not be reserved for those with years or months of sequential, traditional training. Instead, with the right balance of Human mentorship and AI support, these complex skills can quickly become practical and attainable to novices as well. Future work should test 3P across diverse fields, from education and consulting to software engineering, project management, and product development. Future work might also examine how learner profiles and field-specific needs shape 3P's effectiveness, while real-time learning analytics and comparisons of team-based *versus* individual workflows could refine best practices and support scale-up. Longitudinal studies may further assess whether confidence gains persist and translate into lasting professional impact. Cross-disciplinary evaluations could also determine the generalizability of 3P.

As emerging AI approaches expand into complex activities such as deep research, code generation, and project management, their potential to amplify both learning and doing will only grow [49,160–162]. Personalized assistants like Coach AI [147,148] and broader platforms such as ChatGPT Edu [163–166]. may soon meet more nuanced learning needs. Incorporating multimodal, deep research, and agentic AI technologies could further strengthen 3P by enhancing rapid, scalable, and widespread implementation. Similarly to the Cursor-based reanalysis presented here, AI-empowered systems may become key contributors in future 3P implementations [62–65]. Importantly, while AI enables accelerated skill development and result generation, final interpretation, verification, and validation remain Human-driven.

The timeliness of the 3P framework is underscored by two recent reports from Stanford and MIT [167,168]. Stanford's analysis of ADP employment data suggests that entry-level jobs in GenAI-affected fields such as software development and customer services are declining [167]. These and other reports argue that AI is more likely to replace "book-learning" knowledge gained in formal education than experience-based skills, which helps explain stagnant job prospects for younger workers despite overall labor market resilience [167,169,170]. MIT's findings further suggest that most GenAI projects fail not due to technological limitations but because of adoption gaps [168,171]. 3P directly addresses both challenges: it provides a rapid, scalable, field-agnostic blueprint for transforming novices into contributors who generate real-world outcomes. 3P fosters hands-on, Human–AI collaboration to accelerate sophisticated projects, drive innovation, and enable measurable workforce reskilling.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/2511171319045534/LifeAI-1261-Supplemenmtary-Materials.pdf. Table S1: Sample prompts utilized by Team C using GPT-4o OMNI. Table S2: Sample prompts utilized by Teams A and B using Claude 3.7 Sonnet. Figure S1: UnitProt APIs used for initial sets. Figure S2: Comparative analysis of AI prompt performance metrics across the article. Figure S3: Distributions of eukaryotic enzyme lengths. Figure S4: Spectral analysis of eukaryotic enzyme lengths. Figure S5: Statistical mixture modeling of enzyme lengths. Figure S6: Four processed datasets: Results validation with mixture model analysis. Figure details: (A, Top Left) Dataset 1, (B, Top Right) Dataset 2, (C, Bottom Left) Dataset 3, and (D, Bottom Right) Dataset 4.

Author Contributions

N.C., V.S., K.P., I.O., O.R. and E.K. contributed to the conceptualization, methodology, and writing of the manuscript. All the authors (N.C., V.S., K.P., I.O., A.F.Y., B.B.B., C.-H.L., J.B., N.N.A., P.H., R.E.S.D., S.B.N., G.S., N.K., N.S., O.R. and E.K.) contributed to data curation, software development, results validation, and data visualization. N.C., V.S., G.S., N.K., N.S., O.R. and E.K. contributed to the revised version of the manuscript. E.K. provided supervision. All authors have read and agreed to the published version of the manuscript.

Funding

We gratefully acknowledge the support of the MS in Biomedical Informatics program at NYU Langone, led by David Fenyo, for enabling this course. Support for N.K. and E.K. from 1Ekaroni is very much appreciated.

Data Availability Statement

Data produced by this work, including the prompts, initial and processed versions of the four eukaryotic enzyme datasets (Datasets 1–4), and complete analysis outputs, are publicly available. The processed datasets and code used for generating Spectral Analysis of Distributions (SAD), Gamma and Normal Mixture Modeling, likelihood ratio testing (LRT), Fourier analysis, all the tables and figures, and supplementary materials can be found in the GitHub

repository: https://github.com/nyu-vilcek-bmi/BMI_GenAI_Class_2025_LifeAI_Paper (accessed on 14 November 2025). All datasets were originally sourced from the UniProt Knowledgebase (Eukaryota; Taxonomy ID 2759) using API queries and filtering steps detailed in the text and Figures 3, S1, and S3.

Acknowledgments

We sincerely thank Aliona Rudys, David Fenyo, Jose Pagan, Maria Blekher, Oshin Aivazian, Paul Buehrens, Steve Cohen, Jack Faris, Benjamin Kolker, Evelyne Kolker, Eugene Luskin, Dana Mauriello, Nader Mherabi, Craig Muraskin, Sean Snyder, Aleksandr Zhuk, and Dmitrij Frishman for stimulating discussions.

Conflicts of Interest

The authors declare no conflict of interest. Given the role as the Executive Editor, Eugene Kolker had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

Use of AI and AI-Assisted Technologies

This work employed advanced AI and generative AI (GenAI) systems in a controlled, supervised, and fully validated research environment. The tools used include OpenAI's ChatGPT, Anthropic's Claude, Perplexity AI, and Cursor. For instance, ChatGPT (GPT-4o Omni) streamlined structured interactions and documentation; Claude 3.7 Sonnet empowered sophisticated statistical modeling and cross-validation; Perplexity AI delivered authoritative literature search and contextual biological insight; and Cursor enabled collaborative, reproducible code validation. These tools were deployed in a synergistic, interchangeable manner at crucial decision points and at resubmission.

Tool outputs were rigorously reviewed and validated by our team for scientific correctness, technical fidelity, reproducibility, and industry relevance. The interplay of GenAI models (in both isolated and combined workflows) provided actionable insights into optimization opportunities, feature differentiation, and critical integration points between AI and Human domain experts. The authors take full responsibility for the final content.

References

1. Marks, S.E.; Walter, G.A. *Experiential Learning & Change: Theory, Design & Practice*; Wiley: Hoboken, NJ, USA, **1981**.
2. Kolb, D.A. *Experiential Learning: Experience as the Source of Learning and Development*; Financial Times Prentice Hall: Englewood Cliffs, NJ, USA, **1984**.
3. Kolb, D.A. *Experiential Learning: Experience as the Source of Learning and Development*, 2nd ed.; Pearson Education: Upper Saddle River, NJ, USA, **2015**.
4. Black, S.L.; DeGrassi, S.W.; Sweet, K.M. Multisource feedback as an experiential learning enabler in large-format management classes. *J. Manag. Educ.* **2021**, *45*, 479–517. https://doi.org/10.1177/1052562920987292.
5. Cherry, K. The Experiential Learning Theory of David Kolb. Understanding the Four Stages of Learning. **2025**. Available online: https://www.verywellmind.com/experiential-learning-2795154 (accessed on 14 November 2025).
6. U.S. Department of Health Education and Welfare. Public Health Reports. **1953**. Available online: https://babel.hathitrust.org/cgi/pt?id=uiug.30112111844723&seq=255 (accessed on 14 November 2025).
7. Plato 'Apology'. **2025**. Available online: https://classics.mit.edu/Plato/apology.html (accessed on 14 November 2025).
8. Plato, W.P. *The Stanford Encyclopedia of Philosophy (Fall 2024 Edition)*; Zalta, E.N., Nodelman, U., Eds.; **2024**. Available online: https://plato.stanford.edu/archives/fall2024/entries/plato-ethics-shorter/ (accessed on 14 November 2025).
9. Littig, B. The Neo-Socratic Dialogue (NSD): A method of teaching the ethics of sustainable development. In *Teaching Business Sustainability*; Galea, C., Eds.; Greenleaf Publishing: Sheffield, UK, **2004**; Volume 1, pp. 240–252. Available online: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-5415 (accessed on 14 November 2025).
10. Le, N.-T. How Do Technology-Enhanced Learning Tools Support Critical Thinking? *Front. Educ.* **2019**, *4*, 126. Available online: https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2019.00126 (accessed on 14 November 2025).
11. Scholle, C. Understanding the Socratic Method of Teaching. **2020**. Available online: https://www.alu.edu/alublog/understanding-the-socratic-method-of-teaching/ (accessed on 14 November 2025).
12. Giuseffi, F. The investigation of a Nelsonian approach to Socratic dialogue with student-teachers at a midwestern private university. *InSight J. Sch. Teach.* **2024**, *20*, 1. https://doi.org/10.46504/20202401gi.
13. Hood, L.; Heath, J.R.; Phelps, M.E.; et al. Systems biology and new technologies enable predictive and preventative medicine. *Science* **2004**, *306*, 640–643.

14. Galperin, M.Y.; Kolker, E. New metrics for comparative genomics. *Curr. Opin. Biotechnol.* **2006**, *17*, 440–447.
15. Barga, R.; Howe, B.; Beck, D.; et al. Bioinformatics and Data-Intensive Scientific Discovery in the Beginning of the 21st Century. *OMICS J. Integr. Biol.* **2011**, *15*, 199–201.
16. Smith, A.; Balazinska, M.; Baru, C.; et al. Biology and Data-Intensive Scientific Discovery in the Beginning of the 21st Century. *OMICS J. Integr. Biol.* **2011**, *15*, 209–212.
17. Faris, J.; Kolker, E.; Szalay, A.; et al. Communication and data-intensive science in the beginning of the 21st century. *OMICS J. Integr. Biol.* **2011**, *15*, 213–215.
18. Kolker, E.; Stewart, E.; Ozdemir, V. Opportunities and challenges for the life sciences community. *OMICS J. Integr. Biol.* **2012**, *16*, 138–147.
19. Hood, L. Systems Biology and P4 Medicine: Past, Present, and Future. *Rambam Maimonides Med. J.* **2013**, *4*, e0012.
20. Russell, L. Ackoff. General Systems Theory and Systems Research Contrasting Conceptions of Systems Science. In *Views on a General Systems Theory: Proceedings from the Second System Symposium*; Mesarovic, M.D., Ed.; **1968**.
21. Barnes, J. Life and Work. In *The Cambridge Companion to Aristotle*; Cambridge University Press: Cambridge, UK, **1995**.
22. Senge, P.M. *The Fifth Discipline: The Art & Practice of the Learning Organization*; Doubleday: New York, NY, USA, **1990**.
23. Meadows, D.H. *Thinking in Systems: A Primer*; Routledge: Oxfordshire, UK, **2008**.
24. Kolker, E. How Can You Start Using AI? Part 5: When Was the Term 'AI' Coined? **2025**. Available online: https://www.linkedin.com/feed/update/urn:li:activity:7295216336329228288/?updateEntityUrn=urn%3Ali%3Afs_feedUpdate%3A%28V2%2Curn%3Ali%3Aactivity%3A7295216336329228288%29 (accessed on 14 November 2025).
25. Park, S.H. Use of Generative Artificial Intelligence, Including Large Language Models Such as ChatGPT, in Scientific Publications: Policies of KJR and Prominent Authorities. *Korean J. Radiol.* **2023**, *24*, 715–718. https://doi.org/10.3348/kjr.2023.0643.
26. Corbyn, Z. Reid Hoffman: Start using AI deeply. It is a huge intelligence amplifier. **2025**. Available online: https://www.theguardian.com/technology/2025/mar/22/reid-hoffman-superagency-start-using-ai-deeply-it-is-a-huge-intelligence-amplifier (accessed on 14 November 2025).
27. Kolker, E. How Can You Start Using AI? Part 3: Get an AI Coach for YOUR Business. **2025**. Available online: https://www.linkedin.com/posts/eugenekolker_ai-ai-ai-activity-7287560039408492546-UolS/?utm_source=share&utm_medium=member_desktop&rcm=ACoAADdzCHUB17yvhPTBHa3VY-XBx9BbC3A0l7A (accessed on 14 November 2025).
28. Gran, G.; Kolker, E. The Mayfield podcast. Making AI Real.~~: Kolker E.~~ **2025**. Available online: https://www.linkedin.com/posts/makingairealthepodcast_activity-7333592691336720385-IRuK?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAADOL4BtDMMO4K7DZ3pg-dsoRuIIG4w6qY (accessed on 14 November 2025).
29. Introducing ChatGPT. **2022**. Available online: https://openai.com/index/chatgpt/ (accessed on 14 November 2025).
30. Bommasani, R.; Hudson, D.A.; Adeli, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2022**, arXiv.2108.07258. https://doi.org/10.48550/arXiv.2108.07258.
31. Naveed, H.; Khan, A.U.; Qiu, S.; et al. A Comprehensive Overview of Large Language Models. *arXiv* **2024**, arXiv.2307.06435. https://doi.org/10.48550/arXiv.2307.06435.
32. Minaee, S.; Mikolov, T.; Nikzad, N.; et al. Large Language Models: A Survey. *arXiv* **2025**, arXiv.2402.06196. https://doi.org/10.48550/arXiv.2402.06196.
33. Wang, F.; Minhua, L.; Yao, M.; et al. A Survey on Small Language Models in the Era of Large Language Models: Architecture, Capabilities, and Trustworthiness. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 3–7 August 2025. https://doi.org/10.1145/3711896.3736563.
34. Cohen, J.; Rosenfeld, E.; Kolter, J. Z. Certified Adversarial Robustness via Randomized Smoothing. In Proceedings of the 36th International Conference on Machine Learning. *ICML* **2019**, PMLR 97:1310-1320.
35. OpenAI. GPT-4o System Card. *arXiv* **2024**, arXiv:2410.21276. https://arxiv.org/abs/2410.21276.
36. Introducing the next generation of Claude. **2025**. Available online: https://www.anthropic.com/news/claude-3-family (accessed on 14 November 2025).
37. Perplexity. Perplexity AI. Deep Research mode. **2025**. Available online: https://www.perplexity.ai/hub/blog/introducing-perplexity-deep-research (accessed on 14 November 2025).
38. Eger, S.; Cao, Y.; D'Souza, J.; et al. Transforming Science with Large Language Models: A Survey on AI-assisted Scientific Discovery, Experimentation, Content Generation, and Evaluation. *arXiv* **2025**, arXiv.2502.05151. https://doi.org/10.48550/arXiv.2502.05151.
39. Garrity, G.M.; Field, D.; Kyrpides, N.; et al. Toward a standards-compliant genomic and metagenomic publication record. *OMICS J. Integr. Biol.* **2008**, *12*, 157–160.
40. Higdon, R.; van Belle, G.; Kolker, E. A note on the false discovery rate and inconsistent comparisons between experiments. *Bioinformatics* **2008**, *24*, 1225–1228.

41. Louie, B.; Higdon, R.; Kolker, E. A statistical model of protein sequence similarity and function similarity reveals overly-specific function predictions. *PLoS ONE* **2009**, *4*, e7546.
42. Kolker, E.; Higdon, R.; Welch, D.; et al. SPIRE: Systematic protein investigative research environment. *J. Proteom.* **2011**, *75*, 122–126.
43. Dove, E.S.; Faraj, S.A.; Kolker, E. Ozdemir. Designing a post-genomics knowledge ecosystem to translate pharmacogenomics into public health action. *Genome Med.* **2012**, *4*, 91.
44. Ozdemir, V.; Kolker, E.; Hotez, P.J.; et al. Ready to put metadata on the post-2015 development agenda? Linking data publications to responsible innovation and science diplomacy. *OMICS J. Integr. Biol.* **2014**, *18*, 1–9.
45. Higdon, R.; Haynes, W.; Stanberry, L.; et al. Unraveling the complexities of life sciences data. *Big Data* **2015**, *1*, 42–50.
46. Janes, K.A.; Chandran, P.L.; Ford, R.M.; et al. An engineering design approach to systems biology. *OMICS J. Integr Biol.* **2017**, *9*, 574–583. https://doi.org/10.1039/c7ib00014f.
47. Hasin, Y.; Seldin, M.; Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **2017**, *18*, 83. https://doi.org/10.1186/s13059-017-1215-1.
48. Mura, C.; Draizen, E.J.; Bourne, P.E. Structural biology meets data science: Does anything change? *Curr. Opin. Struct. Biol.* **2018**, *52*, 95–102. https://doi.org/10.1016/j.sbi.2018.09.003.
49. Kolker, E. How Can YOU Start Using AI? Part 6: AI Agents. **2025**. Available online: https://www.linkedin.com/feed/update/urn:li:activity:7297627727518486528/ (accessed on 14 November 2025).
50. Frishman D.A. Yet Another AI Journal? *LifeAI* **2025**, *1*, 1. https://www.sciltp.com/journals/lifeai.
51. Chen, L.; Chen, P.; Lin, Z. Artificial Intelligence in Education: A Review. *IEEE Access* **2020**, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510.
52. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention Is All You Need. *arXiv* **2023**, arXiv.1706.03762. https://doi.org/10.48550/arXiv.1706.03762.
53. Wang, S.; Wang, F.; Zhu, Z.; et al. Artificial intelligence in education: A systematic literature review. *Expert Syst. Appl.* **2024**, *252*, 124167. https://doi.org/10.1016/j.eswa.2024.124167.
54. Touvron, H.; Lavril, T.; Izacard, G.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv.2302.13971. https://doi.org/10.48550/arXiv.2302.13971.
55. Buriak, J.M.; Akinwande, D.; Artzi, N.; et al. Best Practices for Using AI When Writing Scientific Manuscripts. *ACS Nano* **2023**, *17*, 4091–4093. https://doi.org/10.1021/acsnano.3c01544.
56. Koller, D.; Beam, A.; Manrai, A.; et al. Why We Support and Encourage the Use of Large Language Models in NEJM AI Submissions. *NEJM AI* **2024**, *1*, AIe2300128. https://doi.org/10.1056/AIe2300128.
57. OpenAI. GPT-4 Technical Report. *arXiv* **2024**, arXiv.2303.08774. https://doi.org/10.48550/arXiv.2303.08774.
58. Möller, M.; Nirmal, G.; Fabietti, D.; et al. Revolutionising Distance Learning: A Comparative Study of Learning Progress With AI-Driven Tutoring. *arXiv* **2024**, arXiv.2403.14642. https://doi.org/10.48550/arXiv.2403.14642.
59. Zhang, J.; Fenton, S.H. Preparing healthcare education for an AI-augmented future. *NPJ Health Syst.* **2024**, *1*, 4. https://doi.org/10.1038/s44401-024-00006-z.
60. Hersh, W. Generative Artificial Intelligence: Implications for Biomedical and Health Professions Education. *Annu. Rev. Biomed. Data Sci.* **2025**, *8*, 355–380. https://doi.org/10.1146/annurev-biodatasci-103123-094756.
61. Zhao, W.X.; Zhou, K.; Li, J.; et al. A Survey of Large Language Models. *arXiv* **2025**, arXiv.2303.18223.
62. Cursor. Best practices for medium/large projects. **2025**. Available online: https://forum.cursor.com/t/best-practices-for-medium-large-projects/21206 (accessed on 14 November 2025).
63. Replit Guides. **2025**. Available online: https://replit.com/guides (accessed on 14 November 2025).
64. Leaderboards. **2025**. Available online: https://www.swebench.com (accessed on 14 November 2025).
65. GPT-5 Codex. **2025**. Available online: https://openai.com/index/introducing-codex/ (accessed on 14 November 2025).
66. Nobel Prize in Chemistry. **2024**. Available online: https://www.nobelprize.org/prizes/chemistry/2024/press-release/ (accessed on 14 November 2025).
67. Krishna, R.; Wang, J.; Ahern, W.; et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **2024**, *384*, eadl2528. https://doi.org/10.1126/science.adl2528.
68. Abramson, J.; Adler, J.; Dunger, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.
69. Kovalevskiy, O.; Mateos-Garcia, J.; Tunyasuvunakool, K. AlphaFold two years on: Validation and impact. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2315002121. https://doi.org/10.1073/pnas.2315002121.
70. Kim, G.B.; Kim, J.Y.; Lee, J.A.; et al. Functional annotation of enzyme-encoding genes using deep learning with transformer layers. *Nat. Commun.* **2023**, *14*, 7370. https://doi.org/10.1038/s41467-023-43216-z.
71. Zheng, L.; Shi, S.; Lu, M.; et al. AnnoPRO: A strategy for protein function annotation based on multi-scale protein representation and a hybrid deep learning of dual-path encoding. *Genome Biol.* **2024**, *25*, 41. https://doi.org/10.1186/s13059-024-03166-1.

72. Berman, A.L.; Kolker, E.; Trifonov, E.N. Underlying order in protein sequence organization. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4044–4047.
73. Welch, W.J.; Eggers, D.K.; Hansen, W.J.; et al. Early Events in the Synthesis and Maturation of Polypeptides. **1999**. Available online: https://www.taylorfrancis.com/chapters/edit/10.1201/9781482283440-19/early-events-synthesis-maturation-polypeptides-welch-eggers-hansen-nagata (accessed on 14 November 2025).
74. Kolker, E.; Tjaden, B.C.; Hubley, R.; et al. Spectral Analysis of Distributions: Finding Periodic Components in Eukaryotic Enzyme Length Data. *OMICS J. Integr. Biol.* **2002**, *6*, 123–130.
75. Lynch, M.; Marinov, G. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15690–15695.
76. Masse, M.M.; Hutchinson, R.B.; Morgan, C.E.; et.al. Mapping Protein–Protein Interactions at Birth: Single-Particle Cryo-EM Analysis of a Ribosome–Nascent Globin Complex. *ACS Cent. Sci.* **2024**, *10*, 385–401.
77. Perin, C.; Cretin, G.; Gelly, J.-C. Hierarchical Analysis of Protein Structures: From Secondary Structures to Protein Units and Domains. In *Protein Supersecondary Structures: Methods and Protocols*; Springer US: New York, NY, USA, **2025**; pp. 357–370. https://doi.org/10.1007/978-1-0716-4213-9_18.
78. Trifonov, E.N. Thirty Years of Multiple Sequence Codes. *Genom. Proteom. Bioinform.* **2011**, *9*, 1–6. https://doi.org/10.1016/S1672-0229(11)60001-6.
79. Trifonov, E.N. Codes of Biosequences. Barbieri, M. (Ed.). *The Codes of LIFE: The Rules of Macroevolution*; Springer: Berlin/Heidelberg, Germany, **2008**.
80. Doolittle, R.F. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **1995**, *64*, 287–314.
81. Bashton, M.; Chothia, C. The geometry of domain combination in proteins. *J. Mol. Biol.* **2002**, *315*, 927–939. https://doi.org/10.1006/jmbi.2001.5288.
82. Chothia, C.; Gough, J.; Vogel, C.; et al. Evolution of the protein repertoire. *Nat. Rev. Genet.* **2003**, *4*, 261–270.
83. Wetlaufer, D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 697–701.
84. Savageau, M.A. Proteins of Escherichia coli come in sizes that are multiples of 14 kDa: Domain concepts & evolutionary implications. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1198–1202.
85. Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; et al. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 778–779.
86. Kelman, Z.; Finkelstein, J.; O'Donnell, M. Protein Structure: Why have six-fold symmetry? *Curr. Biol.* **1995**, *5*, 1239–1242.
87. Kolker, E.; Trifonov, E.N. Sequence sizes of eukaryotic enzymes. In *Pacific Symposium on Biocomputing: Proceedings*; World Scientific: Singapore, **1995**; pp. 351–361.
88. Kolker, E.; Trifonov, E.N. Segments, folds and overall protein structure. *Biol. Struct.Dyn. Proc. Ninth Conversat. Discip. Biomol. Stereodyn.* **1996**, *1*, 257–265.
89. Lindgård, P.-A.; Bohr, H. Magic Numbers in Protein Structures. *Phys. Rev. Lett.* **1996**, *77*, 779–782.
90. Bohr, J.; Bohr, H.; Brunak, S. Protein folding and wring resonances. *Biophys. Chem.* **1997**, *63*, 97–105.
91. Durup, J. On "Levinthal paradox" and the theory of protein folding. *J. Mol. Struct.* **1998**, *424*, 157–169.
92. Xu, D.; Nussinov, R. Favorable domain size in proteins. *Fold. Des.* **1998**, *3*, 11–17. https://doi.org/10.1016/S1359-0278(98)00004-2.
93. Wheelan, S.J.; Marchler-Bauer, A.; Bryant, S.H. Domain size distributions can predict domain boundaries. *Bioinformatics* **2000**, *16*, 613–618. https://doi.org/10.1093/bioinformatics/16.7.613.
94. Dawson, P.E.; Kent, S.B.H. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
95. Baldi, P.; Brunak, S. *Bioinformatics: The Machine Learning Approach*; Dietterich, T., Bishop, C., Heckerman, D., Jordan, M., Kearns, M., Eds.; MIT Press: Cambridge, MA, USA, **2001**.
96. Kimmerlin, T.; Seebach, D. 100 years of peptide synthesis: Ligation methods for peptide and protein synthesis with applications to β-peptide assemblies. *J. Pept. Res.* **2005**, *65*, 229–260.
97. Li, D.J.; Zhang, S. Prediction of genomic properties and classification of life by protein length distributions. *arXiv* **2008**, arXiv:0806.0205.
98. Broglia, R.A. Learning to design resistance proof drugs from folding. *Eur. Phys. J. D* **2009**, *51*, 137–151.
99. Hegazy, U.M.; Mannervik, B. Semisynthetic Enzymes. *Amino Acids Pept. Proteins Org. Chem.* **2009**, *2*, 381–433.
100. Zhao, L.; Lu, W. Mirror image proteins. *Curr. Opin. Chem. Biol.* **2014**, *22*, 56–61.
101. Possenti, A.; Vendruscolo, M.; Camilloni, C.; et al. A method for partitioning the information contained in a protein sequence between its structure and function. *Proteins Struct.* **2018**, *86*, 956–964.
102. Svedberg, T. Mass and size of protein molecules. *Nature* **1929**, *123*, 871.
103. Eisenberg, H. Birth of the macromolecule. *Biophys. Chem.* **1996**, *59*, 247–257.
104. Jordan, I.K.; Rogozin, I.B.; Wolf, Y.I.; et al. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.* **2002**, *12*, 962–968. https://doi.org/10.1101/gr.87702.

105. Zhang J., He X. Significant Impact of Protein Dispensability on the Instantaneous Rate of Protein Evolution. *Mol. Biol. Evol.* **2005**, *22*, 1147–1155. https://doi.org/10.1093/molbev/msi101.
106. Wolf, Y.I.; Gopich, I.V.; Lipman, D.J.; et al. Relative Contributions of Intrinsic Structural–Functional Constraints and Translation Rate to the Evolution of Protein-Coding Genes. *Genome Biol. Evol.* **2010**, *2*, 190–199. https://doi.org/10.1093/gbe/evq010.
107. Dill, K.A. Theory for the Folding and Stability of Globular Proteins. *Biochemistry* **1985**, *24*, 1501–1509.
108. Mahmoud, K.; Poyner, D.P. Protein Chimeras by Exchangeable Modules: A Generic Approach for Engineering Proteins. *Res. J. Cell Mol. Biol.* **2007**, *1*, 1–8.
109. Gerstein, M.; Levitt, M. A structural census of the current population of protein sequences. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11911–11916.
110. Gerstein, M.; Levitt, M. Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci.* **1998**, *7*, 445–456.
111. Apic, G.; Gough, J.; Teichmann, S.A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **2001**, *310*, 311–325. https://doi.org/10.1006/jmbi.2001.4776.
112. Ekman, D.; Björklund, A.K.; Frey-Skött, J.; Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions. *J. Mol. Biol.* **2005**, *348*, 231–243. https://doi.org/10.1016/j.jmb.2005.02.007.
113. Han, J.-H.; Batey, S.; Nickson, A.A.; et al. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 319–330. https://doi.org/10.1038/nrm2144.
114. Kolker, E.; Trifonov, E.N. Periodic recurrence of methionines: Fossil of gene fusion? *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 557–560.
115. Pohlmann, K.C. *Principles of Digital Audio*, 4th ed.; MacGraw-Hill: New York, NY, USA, **2000**.
116. The UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–31.
117. Shapiro, S.; Sanford, S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. https://www.jstor.org/stable/2333709?origin=crossref&seq=1.
118. Irrizarry, A. Introduction to Data Science. Chi-Square Goodness-of-Fit Test. In *Statistical Inference*; **2010**. https://rafalab.dfci.harvard.edu/dsbook/inference.html.
119. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705.
120. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. https://doi.org/10.1214/aos/1176344136.
121. Bayesian Information Criterion—An overview. **2025**. Available online: https://www.sciencedirect.com/topics/medicine-and-dentistry/bayesian-information-criterion (accessed on 14 November 2025).
122. Akaike Information Criterion—An overview. ScienceDirect Topics. **2025**. Available online: https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/akaike-information-criterion (accessed on 14 November 2025).
123. McLachlan, G.J.; Lee, S.X.; Rathnayake, S.I. Finite Mixture Models. **2018**. Available online: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-031017-100325 (accessed on 14 November 2025).
124. Lakshmi, R.V.; Vaidyanathan, V.S. Parameter Estimation in Gamma Mixture Model using Normal-based Approximation. *J. Stat. Theory Appl.* **2016**, *15*, 25–35. https://doi.org/10.2991/jsta.2016.15.1.3.
125. 1.2—Maximum Likelihood Estimation | STAT 415. **2025**. Available online: https://online.stat.psu.edu/stat415/lesson/1/1.2 (accessed on 14 November 2025).
126. Wong, T.S.T.; Li, W.K. Test for homogeneity in gamma mixture models using likelihood ratio. *Comput. Stat. Data Anal.* **2014**, *70*, 127–137. https://doi.org/10.1016/j.csda.2013.09.001.
127. Deeks, J.J.; Altman, D.G. Diagnostic tests 4: Likelihood ratios. *BMJ* **2004**, *329*, 168–169. https://doi.org/10.1136/bmj.329.7458.168.
128. Ebeling, W.; Molgedey, L.; Kurths, J.; et al. Entropy, Complexity, Predictability, and Data Analysis of Time Series and Letter Sequences. In *The Science of Disasters*; Bunde, A., Kropp, J., Schellnhuber, H.J. eds.; Springer: Berlin/Heidelberg, Germany, 1992; pp. 3–47. https://doi.org/10.1093/nar/gkac1052.
129. Strang, G. Wavelet transforms versus Fourier transforms. *Bull. Am. Math. Soc.* **1993**, *28*, 288–305.
130. van der Vaart, A.W. (Ed.). Likelihood Ratio Tests. In Asymptotic Statistics; Cambridge University Press: Cambridge, UK, **1998**; pp. 227–241. https://doi.org/10.1017/CBO9780511802256.017.
131. Muthuswamy, J.; Thakor, N.V. Spectral analysis methods for neurological signals. *J. Neurosci. Methods* **1998**, *83*, 1–14. https://doi.org/10.1016/S0165-0270(98)00065-X.
132. Ahdesmäki, M.; Lähdesmäki, H.; Pearson, R.; et al. Robust detection of periodic time series measured from biological systems. *BMC Bioinform.* **2005**, *6*, 117. https://doi.org/10.1186/1471-2105-6-117.
133. Wang, Z.; Atchley, W. Spectral Analysis of Sequence Variability in Basic-Helix-loop-helix (bHLH) Protein Domains. *Evol. Bioinform.* **2007**, *2*. https://pmc.ncbi.nlm.nih.gov/articles/PMC2674655/.

134. Hatton, L.; Warr, G. Protein structure and evolution: Are they constrained globally by a principle derived from information theory? *PLoS ONE* **2015**, *10*, e0125663. https://doi.org/10.1371/journal.pone.0125663.
135. Jaiswal, S.; Murthy, H.A.; Narayanan, M. SpecGMM: Integrating Spectral analysis and Gaussian Mixture Models for taxonomic classification and identification of discriminative DNA regions. *Bioinforma. Adv.* **2024**, *4*, vbae171. https://doi.org/10.1093/bioadv/vbae171.
136. Fourier Transform An Interactive Guide to the Fourier Transform—Better Explained. **2025**. Available online: https://betterexplained.com/articles/an-interactive-guide-to-the-fourier-transform/ (accessed on 14 November 2025).
137. Short-Time Fourier Transform—An overview. **2025**. Available online: https://www.sciencedirect.com/topics/engineering/short-time-fourier-transform (accessed on 14 November 2025).
138. Exploratory Data Analysis—An overview. **2025**. Available online: https://www.sciencedirect.com/_topics/social-sciences/exploratory-data-analysis?utm_source=chatgpt.com (accessed on 14 November 2025).
139. Kolmogorov-Smirnov Goodness-of-Fit Test. **2025**. Available online: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm (accessed on 14 November 2025).
140. Wright, E.S. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *R J.* **2016**, *8*, 352–359. https://doi.org/10.32614/RJ-2016-025.
141. Wright, E.S. Accurately clustering biological sequences in linear time by relatedness sorting. *Nat. Commun.* **2024**, *15*, 3047. https://doi.org/10.1038/s41467-024-47371-9.
142. Ong, A.Y.; Merle, D.A.; Wanger, S.K.; et al. Exploring the Dilemma of AI Use in Medical Research and Knowledge Synthesis: A Perspective on Deep Research Tools. *J. Med. Internet Res.* **2025**, *27*, e75666.
143. Gilbert, W. Why genes in pieces? *Nature* **1978**, *271*, 501. https://doi.org/10.1038/271501a0.
144. Patthy, L. Exon Shuffling Played a Decisive Role in the Evolution of the Genetic Toolkit for the Multicellular Body Plan of Metazoa. *Genes* **2021**, *12*, 382. https://doi.org/10.3390/genes12030382.
145. Cody, I.A. The Human Element in Information. In *The Palgrave Handbook of Breakthrough Technologies in Contemporary Organisations*; Moussa, M, McMurray, A, Eds.; Springer Nature: Singapore, **2025**; pp. 249–259.
146. Tan, X.; Xu, W.; Wang, C. Voice in AI-assisted multimodal texts: What do readers pay attention to? *Comput. Compos.* **2025**, *75*, 102918. https://doi.org/10.1016/j.compcom.2025.102918.
147. Triola, M.M.; Rodman, A. Integrating Generative Artificial Intelligence into Medical Education: Curriculum, Policy, and Governance Strategies. *Acad. Med. J. Assoc. Am. Med. Coll.* **2025**, *100*, 413–418.
148. Kolker, E. How Can You Start Using AI? Part 11: Meet 'Coach AI' from NYU Grossman. **2025**. Available online: https://www.linkedin.com/pulse/how-can-you-start-using-ai-post-11-meet-coach-from-kolker-gene--rbz8e/ (accessed on 14 November 2025).
149. Guo, H.-H.; Hu, Y.; Liu, K.; et al. FireRedTTS: A Foundation Text-To-Speech Framework for Industry-Level Generative Speech Applications. *arXiv* **2025**, arXiv.2409.03283. https://doi.org/10.48550/arXiv.2409.03283.
150. Introducing OpenAI o3 and o4-Mini. **2025**. Available online: https://openai.com/index/introducing-o3-and-o4-mini/ (accessed on 14 November 2025).
151. Transforming R&D with Agentic AI: Introducing Microsoft Discovery. **2025**. Available online: https://azure.microsoft.com/en-us/blog/transforming-rd-with-agentic-ai-introducing-microsoft-discovery/ (accessed on 14 November 2025).
152. Claude Takes Research to New Places. **2025**. Available online: https://www.anthropic.com/news/research (accessed on 14 November 2025).
153. Introducing Claude 4. **2025**. Available online: https://www.anthropic.com/news/claude-4 (accessed on 14 November 2025).
154. Gemini Deep Researc—Your Personal Research Assistant. Gemini. **2025**. Available online: https://gemini.google/overview/deep-research/?hl=en (accessed on 14 November 2025).
155. Grok. Grok 3. **2025**. Available online: https://www.tryprofound.com/blog/understanding-grok-a-comprehensive-guide-to-grok-websearch-grok-deepsearch (accessed on 14 November 2025).
156. NVIDIA. Deep Research. **2025**. Available online: https://research.nvidia.com/labs/adlr/ (accessed on 14 November 2025).
157. IBM Watsonx Assistant Virtual Agent. **2025**. Available online: https://www.ibm.com/products/watsonx-assistant. https://doi.org/10.1097/ACM.0000000000005963 (accessed on 14 November 2025).
158. OpenAI. Deep Research FAQ. **2025**. Available online: https://help.openai.com/en/articles/10500283-deep-research-faq (accessed on 14 November 2025).
159. OpenAI. GPT-5 System Card. **2025**. Available online: https://cdn.openai.com/gpt-5-system-card.pdf (accessed on 14 November 2025).
160. Georgiou, G.P. Capabilities of GPT-5 across Critical Domains: Is It the Next Breakthrough? *arXiv* **2025**, arXiv:2508.19259. https://arxiv.org/abs/2508.19259.
161. Swanson, K.; Wu, W.; Bulaong, N.L.; et al. The Virtual Lab of AI Agents Designs New SARS-CoV-2 Nanobodies. *Nature* **2025**, *646*, 716–723. https://doi.org/10.1038/s41586-025-09442-9.

162. Gao, S.; Fang, A.; Huang, Y.; et al. Empowering biomedical discovery with AI agents. *Cell* **2025**, *187*, 6125–6151.
163. Introducing ChatGPT Edu. **2024**. Available online: https://openai.com/index/introducing-chatgpt-edu/ (accessed on 14 November 2025).
164. Singer, N. Welcome to Campus. Here's Your ChatGPT. The New York Times. **2025**. Available online: https://www.nytimes.com/2025/06/07/technology/chatgpt-openai-colleges.html?utm_source=grant.beehiiv.com&utm_medium=newsletter&utm_campaign=apple-s-new-operating-system&_bhlid=dd56f94771d120207ff1629e20c381a48e0c8240 (accessed on 14 November 2025).
165. OpenAI. ChatGPT Edu Launch Guide for Higher Ed Universities. OpenAI Academy. **2025**. Available online: https://academy.openai.com/public/resources/step-by-step-launch-guide (accessed on 14 November 2025).
166. Cirra AI. GPT-5: A Technical Analysis of Its Evolution & Features. **2025**. Available online: https://cirra.ai/articles/pdfs/gpt-5-technical-overview.pdf (accessed on 14 November 2025).
167. Brynjolfsson, E.; Chandar, B.; Chen, R. Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. **2025**. Available online: https://digitaleconomy.stanford.edu/wp-content/uploads/2025/08/Canaries_BrynjolfssonChandarChen.pdf (accessed on 14 November 2025).
168. Challapally, A.; Pease, C.; Raskar, R.; et al. The GenAI Divide: State of AI in Business. **2025**. Available online: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf (accessed on 14 November 2025).
169. Lichtenberg, N. First-of-Its-Kind Stanford Study Says AI Is Starting to Have a 'Significant and Disproportionate Impact' on Entry-Level Workers in the U.S. Fortune. **2025**. Available online: https://fortune.com/2025/08/26/stanford-ai-entry-level-jobs-gen-z-erik-brynjolfsson/ (accessed on 14 November 2025).
170. Saini, K. Young Software Developers Losing Jobs to AI, Stanford Study Confirms. Final Round. **2025**. Available online: https://www.finalroundai.com/blog/stanford-study-shows-young-software-developers-losing-jobs-to-ai (accessed on 14 November 2025).
171. Estrada, S. MIT Report: 95% of Generative AI Pilots at Companies Are Failing. Fortune. **2025**. Available online: https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/ (accessed on 14 November 2025).