
RESEARCH JANUARY 28, 2026

HOW PROFESSION SHAPES LLM USAGE: INSIGHTS FROM SEAL SHOWDOWN

By Janie Gu, Jaehwan Jeong, David Lee, Bing Liu, Zihao Wang

AI is rapidly changing the nature of work for almost every single career path. In this study, we asked a simple but critical question: how much a user's profession shapes their LLM usage, and in what ways? Using over 580k+ real production prompts and 100k+ side-by-side model battles from SEAL Showdown, we found that profession, not just topic, strongly shapes how people use LLMs, what they ask, and which models they prefer. Some interesting findings:

- **Usage diverges by profession:** STEM users primarily use LLMs for problem solving; Legal professionals heavily use LLMs for drafting and refinement.
- **Experts ask harder questions:** Legal and medical professionals ask a much higher share of high-difficulty, in-domain prompts than non-experts.
- **Same domain, different needs:** Doctors ask about clinical treatment and patient care; non-experts ask about fitness and self-diagnosis. Engineers ask about infrastructure, APIs, and security; non-experts ask more about AI models and data tools.
- **Model preferences differ:** Profession significantly affects model rankings. For coding tasks, software engineers and non-experts prefer different model families.

We hope these insights can help the community build better models not just for everyday users, but for professional use cases.

What is SEAL Showdown?

SEAL Showdown is a ranking of large language models (LLMs) designed to reflect human preferences in a natural chat setting. Scale's contributors have access to an internal, LLM-API-agnostic chat application that gives them free access to all frontier models for their everyday workflows (personal and professional).

During conversations, users are periodically prompted to participate in a side-by-side model comparison, which is also known as a "battle." Voting is entirely optional: annotators can opt out altogether, and even when asked, they can always skip.

About the Data

We looked at both battle prompts (side-by-side model comparisons) and non-battle prompts (single model responses) to understand differences in user preferences for LLMs and their usage behavior.

For the non-battle prompts, we analyzed 583,043 first-turn user prompts from our chat application submitted between Oct 1, 2024 and Sep 30, 2025. We first extracted the users' most recent job title from their resume and used an LLM classifier to map the title to one of 12 pre-defined profession categories.

For the battle prompts, we analyzed 108,176 user prompts from SEAL Showdown submitted between May 1, 2024 and Nov 30, 2025. We followed the pipeline above to extract the user professions.

What We Learned

Different Professions Use LLMs Differently

We identified four high-level use cases for interacting with LLMs and observed how the user behavior varies based on profession.

[Table: see online version]

Analysis of LLM usage across all users shows that Problem Solving dominates usage, followed by Knowledge, Language & Composition, and ChitChat, which makes up less than 5% of the total usage.

However, we observed a few differences by profession:

- For Legal & Policy professionals, Language & Composition (32%) surpasses Knowledge (29%) as the second most frequent use case, suggesting that legal professionals often use LLMs for drafting, summarizing, and refining complex textual material.
- Conversely, STEM fields—including Software Engineering (71%), Data, IT & Cybersecurity (58%), and Engineering & Science (52%)—exhibit a particularly high reliance on LLMs for Problem Solving.

We then classified the conversations into different topic categories using the following taxonomy:

[Table: see online version]

Then, we identified the three most frequent topics that professionals were chatting about, by use case and profession. The three professions we focused on were: Healthcare & Medicine, Legal & Policy, Software Engineering.

We observed a strong correlation between a user's profession and the topics they chat with LLMs about:

- We see a high number of Health & Medicine (34.5%) topics from the Knowledge type prompts from healthcare professionals.
- Similarly, we see a lot of Technology & Software Assistance prompts (77.5%) from the Problem Solving usage type from software engineers and Legal Assistance & Compliance (32.8%) from the Knowledge usage type from law professionals.
- For legal experts, Legal Assistance & Compliance (42.8%) ranks highest for the Problem Solving category and Refinement (38.4%) for the Language and Composition category.

Correlation Between Prompt Domain Difficulty and Expertise

We examined whether a user's domain of expertise influences the difficulty of the prompts they ask. We hypothesized that domain experts, such as doctors, may ask about rare medical phenomena, while everyday users may ask more general questions, like how to self-diagnose a common flu. To test this, we analyzed how the difficulty distribution varies with user expertise. We use the difficulty definitions provided in the [tech report](#). Given the low occurrence of "Ultra High" difficulty prompts, the subsequent analysis focuses on the four difficulty types: "Ultra Low," "Low," "Medium," and "High."

We observe that domain experts across all professions ask more difficult questions, particularly in the medical and legal fields. Over half of the in-domain prompts from legal and medical professionals were classified as high-difficulty (50.7% and 61.2%, respectively). While Software Engineering and Data, IT & Cybersecurity also rank highly for difficult prompts concerning Technology & Software Assistance (first and third, respectively), the difference in difficulty between expert and non-expert groups is less pronounced than in the medical and legal professions.

Deep Dive on Professional LLM Usage

In addition to analyzing professional LLM usage based on topic and domain difficulty, we conducted clustering analysis to surface common topics that are more likely to be asked by a professional than an average consumer (non-professional) and vice versa.

Healthcare & Medicine

We looked at the questions people ask AI chatbots about Health & Medicine. We found a clear difference between what healthcare professionals (experts) and the general public (non-professionals) ask.

What Healthcare Experts Ask:

When healthcare professionals use AI chatbots, they focus on topics directly related to their jobs, such as patient care and complex medical tasks.

- They are much more likely to ask about clinical treatment for serious conditions, such as severe psychiatric disorders or surgical operations (24.16% and 23.77% of these questions come from experts, compared to the overall average of 11.02% for all health questions).
- Other common expert topics include managing digital patient records and direct patient care protocols (like nursing interventions).

[Table: see online version]

What the General Public (Non-Experts) Ask:

In contrast, non-professionals mainly use AI chatbots for questions about their personal health, self-diagnosis, and general well-being.

They are far more likely to ask about **improving personal health**, such as exercise routines, sleep, fatigue, or common skin conditions.

- For example, questions about "Biological mechanisms, management strategies, and causes of sleep and fatigue" are rarely asked by professionals (only 2.54% of these questions are from experts, compared to the overall average of 11.02% for all health questions).

[Table: see online version]

Technology

We looked at the questions people ask AI chatbots about technology and software help. We found a clear difference between what professionals (Software Engineers and Data/IT Experts) and the general public (non-professionals) ask.

What Tech Experts Ask:

When tech professionals use AI chatbots, they focus heavily on the practical side of building, deploying, and managing software in a live, corporate environment.

- They are most likely to ask about setting up, securing, and managing large-scale infrastructure like containers and Kubernetes clusters (50.00% of these questions come from experts, compared to the overall average of 30.40% for all technology questions).
- Other common expert topics include creating documentation and blueprints for software systems (APIs) and handling encryption/security keys.

[Table: see online version]

Legal & Policy

What Legal Experts Ask:

Legal and policy experts primarily use AI chatbots as a tool to help them with the detailed work of their actual job—specifically, drafting and reviewing legal documents.

- They are more likely than average to ask the chatbot for help with writing pleadings, motions, and briefs (16.58% of these questions are from experts, compared to the overall average of 3.73% for all legal & policy questions).
- Other common expert-led questions involve writing and structuring commercial agreements and clarifying the meaning of contract terms.

[Table: see online version]

What the General Public (Non-Experts) Ask:

We found one major area where the general public asks for help more often than legal experts: immigration.

- Questions about regulations on citizenship, visas, and border control are commonly asked by non-professionals. Only a small fraction (1.52%) of these questions come from legal experts, compared to the overall average of 3.73% for all legal & policy questions.

[Table: see online version]

Model Performance Differs Across Professions

Here we narrow the scope of our analysis to the side-by-side battle results collected through SEAL Showdown and see if battles collected from different professions generate different Elo rankings. To see if different professionals truly have different preferences, we conducted likelihood ratio tests comparing nested logistic regression models. We tested whether adding profession features improves model fit compared to a baseline model that only includes style control features (see [tech report](#) for more detail on style control features).

We found that profession significantly affects preferences (LR = 392.62, df = 209, p < 0.001). This effect persisted even when we controlled for the prompt's topic as well (LR = 304.65, df = 209, p < 0.001), suggesting that a user's profession affects their model preferences beyond broad topic categories.

Check out the [SEAL Showdown](#) website to access the live Elo rankings for different professions.

Case Study: Software Engineers' LLM Preferences for Coding Tasks Versus the Average Consumer

Here we did a deep dive into one cross section of user profession and prompt topic. In particular, we analyzed 32,535 battles collected before Nov 30, 2025 that were about a coding task and compared the Elo rankings generated by coding battles collected from experts (users with "Software Engineering" profession) and non-experts (other professionals).

Below are the model performance judged by two groups of users. We can see experts have different preferences from non-experts. For example, engineering experts seem to prefer models from Claude 4.5 family more than non-experts as we see a huge bump in the elo scores for all four models of the Claude 4.5 family: `claude-sonnet-4-5-20250929`, `claude-sonnet-4-5-20250929 (Thinking)`, `claude-haiku-4-5-20251001`, and `claude-haiku-4-5-20251001 (Thinking)` between non-expert and expert.

* Elo scores are scaled with using `llama4-maverick-instruct-basic` as the reference model

What's Next

We hope these insights help the community build better models not just for everyday users, but for professional use cases. As LLMs move deeper into expert workflows, profession-aware evaluation, training, and product design will become increasingly important. We invite the community to use these findings to inform model development, alignment strategies, and future benchmarks that reflect how LLMs are actually used in the real world.