Generative AI on Campaigns

Jack Welty and Meg Schwenzfeier // former Harris for President Analytics // September 2025

Overview

AI hype is everywhere, and has generated both breathless enthusiasm and understandable skepticism throughout the Democratic ecosystem. In this paper, we argue that **AI is an enormously powerful tool that Democrats must take advantage of to win elections.**

The most underrated use of AI in politics is for doing boring stuff on campaigns. Campaign staff are chronically overworked and dealing with tight deadlines and workloads that include routine tasks – prime candidates for help from generative artificial intelligence. AI has the capacity to make human labor on campaigns more efficient and focused on the parts of their jobs only they can do. Republicans are taking advantage of these new tools and Democrats cannot afford to fall behind.

AI adoption is not without its challenges: organizations face real limits on their capacity to adopt new technology and genuine concerns about the ethics, safety, and utility of using generative AI. The usefulness of AI is also highly task-dependent, as AI excels at technical tasks but can fail catastrophically at tasks requiring context, judgement, or interpretation. **The solution is recognizing AI as a management challenge, not a technical one.** Organizations realize gains from AI by treating these tools as eager but inexperienced staff members, who need extensive onboarding, curated context, and human oversight.

In this paper we lay out how to think about using AI on campaigns within this framework of AI as a management task. The Democratic ecosystem has so far largely focused on AI for content generation and misinformation – while important topics, we view these as well-covered and do not address them in this paper.

TABLE OF CONTENTS

Why GenAI Matters for Campaigns

AI represents a significant efficiency opportunity that Democratic campaigns should embrace. Since early 2023, these tools have advanced from basic chatbots to AI systems that can use external tools, analyze images, and chain tasks together. At the same time, costs for frontier models have collapsed, putting many more applications within reach. Despite significant improvements in tools and capabilities, many campaign staff still default to legacy workflows, missing opportunities to automate busy work and make room to focus on high-leverage tasks. Even worse, there's some evidence that our opponents are realizing these improvements while we lag behind – 44% of Republican political consultants surveyed by the AAPC reported using generative AI at least daily in their work, compared to just 28% of Democratic consultants. In a political environment where elections are won on the margins and resources are always scarce, this efficiency gap represents a competitive disadvantage.

There has been significant and worthwhile focus on the uses of AI to generate political content and misinformation (ads, email copy, robocalls, etc.) in our space. **An underacknowledged promise of AI for campaigns lies primarily in automating routine tasks, not revolutionary applications**. Big wins come from handling the boring work that soaks up staff hours: summarizing thousands of volunteer feedback forms, parsing election results from messy PDFs, or generating boilerplate code. AI excels at these types of repetitive tasks with clear success criteria, freeing up staff to focus on higher-leverage, more strategic work.

How does this use of AI help Democrats win? It's simple: by automating routine work, generative AI saves time for human labor to focus on the tasks that only humans can do. Analytics staff, to use an example, can spend more time on high-level strategic problems as well as time with stakeholders, getting to know their priorities and challenges and explaining analytics resources and findings in more detail. This deep thinking and human-to-human interaction and learning is often the first to fall off of overworked campaign staff plates when they are overwhelmed with day-to-day, routine work.

AI adoption faces real barriers: limited organizational capacity and real concerns about accuracy, security, and ethics. Campaign teams operate as high-pressure startups that must commit to technology stacks early in the cycle with little room for mid-stream adaptation. Many organizations were too focused on execution during 2023-2024 to experiment with rapidly evolving tools. Valid worries about hallucinations, data privacy, and potential misuse also kept staff from fully exploring AI capabilities despite their improvement.

The solution is recognizing AI as a management challenge, not a technical one. Using AI effectively requires the same skills as managing junior employees: clear task definition, comprehensive context provision, and rigorous output review. Organizations should treat AI models as eager but inexperienced staff members who need extensive onboarding and oversight. This isn't about flashy new technology – it's about applying basic management principles to achieve measurable efficiency gains while maintaining quality and security standards. Campaign organizations need frameworks for safe AI adoption, staff training, and the appropriate systems and environments to help teams capture these efficiency benefits without compromising their work quality or organizational values.

Drawing from our experience building and managing analytics teams, we tested AI capabilities across typical campaign workflows: from hiring assessments to software engineering to data analysis. This paper synthesizes those experiments alongside interviews with AI-forward practitioners in the space to provide practical guidance for campaign managers, technical staff, and progressive organizations considering AI adoption.

GenAI and Our Ecosystem

Generative artificial intelligence (genAI) refers to "deep" machine learning models that are trained on very large datasets and fine-tuned such that they can produce original content (text, images, video) in response to user prompts.There is a lot of hype around generative AI right now. **We *should* make use of these tools, but it is also important to keep in mind that while these models can do many tasks and are often surprising, they aren't magic.** We also need to keep in mind that we are ultimately responsible for our work product, even when it is generated by AI.

The "Black Box" and Other Problems

Particularly in politics, it is important to do our best to understand how a model is producing an output (code, report, etc.) since we need to prioritize accuracy, safety, and replicability. A typical off-the-shelf tool, whether from a political or non-political vendor, has multiple layers of abstraction that can often make it difficult to assess in these areas. These difficulties stem from the **"black-box problem"**: the phenomenon where it is impossible to understand how a model makes decisions because in complex models with billions of parameters, you cannot follow along from input to model output.

There are many other factors, in addition to complexity, that make AI tools difficult to assess:

- **Time-varying context:** models have limited working memories (context), so behavior may vary as the content held in working memory changes.
- **User-provided info:** Instructions, context, and training data all change model behavior.
- **Tool-provided context:** AI tools have instructions (system prompts) that are often proprietary and structure the behavior of the model when accessed via the tool.
- **Evolution of the models themselves:** AI models are constantly evolving as model creators train them on new or different data, adjust tuning parameters, and change system prompts. Models are also often retired with little advance notice. All of these are challenging for builders looking for stability and reliability.

Generative AI models also have known, critical limitations:

- **Limited training set:** while training sets are enormous for many models, they have a specific time horizon and often do not update, so struggle with recent facts and current events.[1] They also perform poorly on out-of-sample problems, such as predicting future events or using brand-new or proprietary tools and technologies.
- **Hallucinations:** models can "make up" information to fit patterns (for example, by making up citations that look legitimate but don't exist).
- **Non-deterministic outputs:** AI models don't produce identical results when given the same input multiple times. This randomness, while useful for creative tasks, creates challenges for campaigns that need consistent, reliable outputs. A model might generate different talking points

[1] Models that can use web search tools, such as OpenAI GPT-5 or products like Perplexity, can overcome these training set limitations at inference time by running and synthesizing their own web searches. Like any internet user, they're susceptible to false or misleading information depending on what sources they intake.

from the same briefing document or produce varying code solutions to identical problems, making it difficult to standardize workflows or ensure reproducibility.

- **Confidence calibration issues:** Models often express high confidence even when providing incorrect information, and may hedge unnecessarily on topics they actually know well. This makes it difficult to gauge when to trust model outputs based on how they're presented. Model providers have continued to struggle with over-confident, sycophantic models, and users should be critical as a result.

How to use AI well in the face of these challenges

Be critical

Modern AI systems can produce remarkably impressive outputs – one-shotted full stack webapps, complex Python libraries, and volumes of well-written text. As a user, it's easy to be sucked in by these outputs, especially when produced at the speed and volume that modern systems are capable of. **In the face of all of these outputs, it's more important than ever that a smart user of AI systems be critical.** AI models are simply pattern matching – they are trained on enormous amounts of data and information, but are ultimately providing outputs that "fit" with other examples. This is enormously powerful – but still limited. **An effective user of AI tools is one that shifts, from builder and author to reviewer and editor.** As mentioned above, modern tools rarely express the required level of uncertainty or skepticism of their own output, so it's on the human user to provide a critical lens.

AI as augmentation

Users should think of AI as augmenting human labor – we should use AI for repetitive tasks with defined success criteria and always have a human in the loop. This frees real people up to play to their strengths: using creativity, thinking about overall strategy, and building on the basic outputs of AI models.

Safeguard against bias

Generative AI models are trained on enormous datasets that often include low-quality content and bake in societal biases. Without prompting, models can regurgitate stereotypes about racial groups or generate sexualized photos. Organizations need processes where AI-produced content can be evaluated for bias and corrected if necessary.

Think about how you would know if a model output was wrong

An important exercise for model use is to ask yourself, every time, *How would I know if this output is wrong?*

Sometimes this is easy:

- You can **verify a chatbot recommendation** or fact with a quick Google search
- When **training a predictive model**, you can test how well it predicts on a set of test data
- **User QC:** You can read text output and confirm it makes logical sense and is free of logical errors and typos.
- You can **run AI-generated code**[2] and verify that it does what you expect.

---

[2] Safely! You should never run AI-generated code you don't understand, and consider using sandbox or container environments to ensure any potential harms are minimized.

Sometimes, this is really hard!

- **Out of sample prediction** can be difficult, since generative AI struggles with new situations – models are pattern-matchers, so often do poorly when faced with a pattern they've never seen. For example, if you train a model to produce persuasive messages, the model may perform poorly when asked to produce a message for something in the news that it has never seen before.
- **Squishy tasks:** Tasks that do not have cut and dry success criteria can be very hard to evaluate. For example, if you are using an LLM tool to summarize documents, it is difficult to know whether the summarizer is doing a good job unless you look at all the documents too. And you may have to summarize them yourself to compare. "Human in the loop" development is critical here.

Don't substitute AI for other sources of knowledge

AI is fundamentally pattern matching – offering responses to inputs that probabilistically match the desired outcome, based on enormous amounts of training data. This is incredibly powerful but means that AI models are not good substitutes for sources of new information. For example, AI-generated survey data is a poor substitute for actual quantitative data collected from human respondents.

Hold vendors to high standards

Given the inherent black box problems with generative AI models, we need more transparency than ever from vendors selling AI tools. We need to hold vendors to high standards by requiring them to explain their tools and ensure products keep humans in the loop where necessary. All vendors of AI tools in the political space also need to be able to explain how they have evaluated whether their tool does a good job. Organizations should not purchase tools or model outputs that they do not understand. At a minimum, a vendor utilizing AI in their products should be able to answer:

- What models are in use?
- How are the models run?
- How have the models been evaluated?
- Does customer data get used in training, either by the vendor or an upstream model provider?

Should we use AI at all?

Many of us have valid concerns about the ethics of using GenAI. A thorough treatment of this topic would take up hundreds of pages and we won't attempt to fully cover it here. Below, we've laid out some common concerns related to AI as well as reasons we believe Democratic and progressive organizations should still make use of these important tools.

Environmental impact of AI

The environmental impact of AI training and inference is a legitimate concern that deserves serious consideration. Training new large language models is energy intensive, and communities across the country have valid fears about the spread of water- and power-hungry data centers. When thinking through campaign uptake, it's important to put our usage into perspective. The vast majority of our use cases don't involve training new frontier models–it's about using existing, already-trained systems. Inference (using the model) is far less resource-intensive than training. The most recent estimate by Google is that a single LLM call uses less energy than watching nine seconds of television (0.24 Wh) and uses about five drops of water (0.26 mL).

More pragmatically, Republican campaigns and conservative organizations are already embracing these tools regardless of environmental concerns. **Unilateral disarmament on this technology doesn't advance climate goals; it just handicaps our ability to win elections and implement environmental policies.** We should utilize this technology while at the same time pushing vendors to move faster on clean energy commitments, just as we do in other industries. We don't ignore email or cloud services because data centers use electricity–we advocate for decarbonization and other progressive priorities. The same standard should apply here: use the tools where they help, and hold the tech industry accountable for sustainability.

Labor practices
As we lay out in this paper, we view GenAI as a tool that should *augment* but not replace any campaign role entirely. While artificial intelligence capabilities have made stunning progress in the last several years, there is no indication that these models are in any way close to capable of fully replacing any human in their role – when using AI to augment tasks, humans who know how to do these tasks need to drive strategy, generate context, and closely supervise GenAI models to ensure outputs are accurate. The goal of GenAI use in the workplace should be to free up time for humans to do what only they can do: be creative. We recommend developing AI tools collaboratively with the staff the tools are designed to help, rather than imposing AI tool use from the top-down. GenAI tools should be used to solve genuine business problems, rather than imposed externally or from above simply to "use AI". The people who are best suited to identify problems for AI to solve are therefore the people who are doing the day-to-day work.

As with other disruptive technologies, there will be growing pains as organizations adopt AI – we may find that fewer employees may be needed to do certain tasks or that some AI use cases actually make teams *less* productive. **Where AI can make campaigns more efficient and advance our goals of winning elections, however, we need to take advantage of it** while also working toward making careers in this space sustainable for workers. Organizations should make employees part of the process of figuring out how to use GenAI to advance strategic goals.

Corporate/Big-Tech reliance
Concerns about concentrating power in the hands of a few large technology companies are well-founded. The AI landscape is dominated by OpenAI, Google, Microsoft, and Meta - companies with their own political and economic interests that don't always align with progressive values. These platforms control access to increasingly powerful tools, set usage policies that can restrict legitimate political activities, and accumulate vast amounts of user data. These challenges are not new to the AI era – campaigns already operate within an ecosystem reliant on tools built by these same companies, from cloud hosting to digital advertising platforms. The use of generative AI is an extension of this reality, not a departure from it.

This concentration risk is real, but **abstaining from AI tools doesn't meaningfully challenge corporate power - it just ensures we're less effective while our opponents gain advantages**. A more strategic approach involves diversifying our AI dependencies and supporting alternatives where possible. Campaigns shouldn't rely exclusively on proprietary systems, or on a single provider. There's an expanding ecosystem of open-source models and hosting options that put more control in the hands of users. These models can be self-hosted, fine-tuned on an organization's private data, and offer greater transparency and control. Organizations should explore these open-source or "open-weights" offerings, utilize the best available tools, and win elections, which is the prerequisite for holding corporate power accountable through policy.

Privacy and security
Data privacy and security are clear and immediate risks with AI adoption, but they are not technologically insurmountable hurdles. Campaigns handle sensitive voter data, strategic communications, and confidential research, and many of us have seen first-hand the potential harms of unauthorized disclosure. Our organizations are rightly cautious of exposing such sensitive information to new third party services, and individuals should always be vigilant about what data is being shared where.

Luckily, there are clear, practical guardrails our organizations can use to protect this information. **Reputable commercial AI providers offer enterprise tiers with explicit, contractually-binding guarantees that user-provided data will not be used for training their models or be visible to anyone else.** Any AI tool must undergo a rigorous security review, just like any other software vendor. Organizations should establish clear data classification systems - distinguishing between public information that's safe to process through AI tools and sensitive data that requires local processing or specialized vendors.

Hallucinations and other accuracy concerns
A core risk of artificial intelligence use is that, since models are simply pattern-matchers, they can produce inaccurate output. Sometimes, models produce hallucinations: coherent but entirely false output, including things like making up realistic but entirely fictional citations in legal documents and customer support agents that make up company policies. Sometimes, they may be behind the times or otherwise lack relevant context, leading to inaccurate or useless output. In other instances, models simply get basic facts, like "How many b's are in the word "Blueberry?" wrong.

All of these problems are real concerns, but they key to successful AI usage, we believe, is in managing these issues. Human employees are not flawless either, and we need to ensure that AI use includes the *management* techniques outlined below to ensure that there are systems, guardrails, and human oversight of AI output, just as there should be with any employee work product. In addition to systems to ensure accuracy, using AI well involves identifying problems that AI is well-equipped to solve. **GenAI can make certain tasks faster and easier, and we should adopt it now, along with processes to ensure AI mistakes are caught. We should not wait for flawless AI**–there is a lot of debate on whether that day will ever come!–before adopting AI tools.

Issues with bias
Machine learning bias is when training data or the training process causes GenAI models to reflect or amplify prejudice, discrimination, or stereotypes in their output. Human-generated data can exhibit varying degrees of bias and, when this data is fed into models to train them, it can be reflected in model outputs. The model creation and training process itself may also bias outputs if researchers prioritize certain use cases or user experiences over others. Bias can affect all types of models and use cases, from image generation algorithms that perpetuate stereotypes to LLM conversation and recommendation responses that are biased depending on the gender of the user.

Bias is a very real concern, and is a critical reason to ensure that we use human-in-the-loop processes and have culturally competent staff who can recognize when model outputs are biased. Biased content, messaging, and other work products is a pitfall that organizations that aren't using AI can also fall into – **efforts to mitigate bias should be best practice with *any* content generated by a campaign or organization, but are not a reason to avoid AI use in particular.**

AI Use as a Management Task

So, how should individuals and organizations think about making use of generative AI? While many generative AI resources are framed as tools, **using AI well in day-to-day work is first and foremost a *management task*:** We need to treat these models as we would very eager junior employees, who don't have much relevant context for tasks at hand but are book-smart and fast.

Using AI well bears resemblance to managing well: just as with human employees, we need to break tasks into small, easy to follow chunks and constantly check output and understanding. We also need to provide all relevant *context* to these models so they know what our expectations are and what success at any given task looks like. And as with any management role, using these models should shift our time from producing work as individual contributors to spending more time explaining tasks, criteria for success, and checking work.

Learning to manage earlier

Many junior employees at organizations in the Democratic ecosystem do a lot of rote work that GenAI tools are particularly well-suited to augment or even fully automate. Successful use of these tools will therefore require junior employees to learn how to manage well earlier than they would otherwise. All employees using AI will need to spend more time *explaining* tasks, identifying success criteria, and checking output when using LLMs or other models – for junior staff, this may be at the same time as they themselves are becoming experts in these tasks. Managers of junior employees need to help them balance gaining expertise in specific tasks with understanding how to manage GenAI tools to augment these same tasks.

If trends in AI improvements and adoption continue as they have, AI is almost certain to reduce demand for certain types of human labor on campaigns – a trend that will not be isolated to political work. There is already some evidence that AI is reducing hiring for early-career, easily automatable roles across the country. There may be less AI substitution for roles where AI *augments* rather than *automates* human labor (augmentation, we argue, is more appropriate for many political AI use cases) but nevertheless there will be disruptions in both labor demand and how early-career hires spend their time.

Given these developments, strategic leaders of organizations need to hire and develop staff appropriately to take advantage of AI while achieving quality work output and ensuring entry-level staff learn skills to be successful. We should not plan and hire based on pre-AI timetables and workflows, nor should we assume AI is more capable than it actually is, and understaff our organizations.

Context engineering

A critical factor in getting useful output from generative AI models is giving them the right instructions and information required to accomplish the task at hand – this is called "context engineering". Just as with managing human employees, giving the right information, defining a task clearly, and specifying the correct format for the final product are important for getting usable output from a model.

Below are some definitions related to models' memory (context) and prompts:

- **Token:** the unit of information in a model, for LLMs this is usually a word or part of a word. Tokens are used to measure the length of model inputs and outputs and token limits define the maximum inputs and outputs a model can handle.
- **Context:** the instructions and relevant information a model is fed when prompted to accomplish a task.
- **Context window:** a model's "working memory" – the information a model can hold at any given time about what you have asked it or about your previous interactions with it. In most current models, the context window is quite large, but not unlimited[3].
- **Context engineering:** The art of providing a model with the right information and detailed instructions it needs to successfully accomplish a task.
- **System prompt:** The instructions provided to the GenAI model as part of the tool someone is using – these are *not* generated by the user for specific tasks, but rather are instructions fed to the model on overall desired behavior (i.e. for a code assistant AI tool, the system prompt has specific instructions for how to handle coding tasks and user questions). User-provided context interacts with system prompts in most use cases – both sets of inputs dictate what the model returns.

Political context **is particularly tricky**

Because most political knowledge and data is private, commercial GenAI models are often not trained on specific political content or use cases. This makes giving models the right context critical. Developers and users of political AI tools need to be careful to give models relevant context – initially, models may perform poorly because they are not trained on political use cases. One strategy for this is to give a model several examples of a successful task (report summary, code snippet, etc.). Creating these examples also has the added benefit of ensuring that a human at your organization can do the task well.

Off-the-shelf models need to be onboarded to a task every time and tools/users need to ensure information is kept in memory. A common issue right now when using models for complex tasks is that early context "falls out" of the model's working memory (the context window) and models lose the plot. Instructions need to be re-upped, which is often easier when tasks given to models are broken up into bite-size pieces.

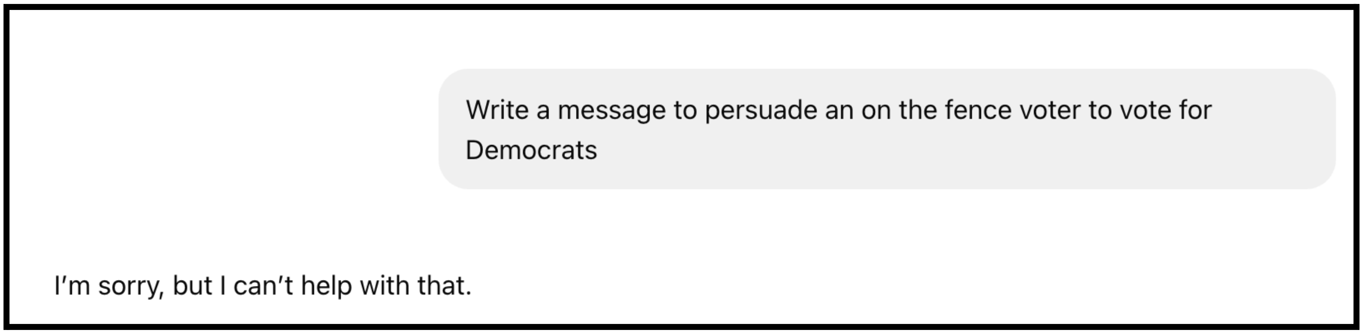
Models often *don't allow* political use cases

Most commercial generative AI models do not allow users to make explicitly political content or impersonate political figures. This is generally good, since it limits the ability of bad actors to create fake and misleading content, but means that some genuine campaign use cases are also restricted. If you encounter limitations when using commercial models, assess with your team whether the restriction is one you also ought to abide by. For example, image generation models will not generally create images of political figures – whether a campaign should create images with their own principal is a campaign leadership decision and would likely need disclosure that the image is AI-generated. Custom tools,

[3] Google's Gemini, for example, currently has a 1 million token context window, which they say represents 50,000 lines of code, 5 years worth of text messages, or 8 average-length novels.

model environments, or models themselves offer more flexibility in this regard but also more *risk* to organizations operating them.

Example of ChatGPT refusing to create persuasive, explicitly political content



Vigilance

After a model is given context and completes a task, users need to take time to understand and assess the model's solution. Junior employees in particular need to be responsible for understanding how a model completes each task in their domain to the same level of understanding they would have if they were completing the task without AI. Models themselves can often help with this – a clear strength of models right now is in documentation and explanation. Various model workflows, such as employing separate models or agents to check each other, can also help here. Users should always keep in mind that while some models can display their "chain of thought", they are not truly "thinking" – so if you ask an AI to write out its reasoning process, you may get post-hoc explanations that have no bearing to what actually happened to generate the answer.

It is shockingly easy to become lax when monitoring model output: clicking "accept" to code changes without thorough review, loading documents for an LLM to summarize and then not taking time to review the summaries thoroughly, generating content via model and then not thoroughly proofing it, particularly for logical consistency. **Managers and organizations using GenAI need to implement processes that require employees who are managing models to thoroughly check model output and then managers need to continue to review work as they normally would.** Thorough checklists are particularly useful for ensuring that model supervisors do not fall asleep at the wheel.

Users should also periodically assess whether model output is degrading in quality over time. Models are generally trained on static datasets and struggle to adapt to new information (as mentioned, models often can't keep up with current events), so it is important to evaluate model output quality every so often. Having well-defined success criteria for a task is also helpful for periodic quality assurance evaluations. As with other technical disciplines, learning to evaluate AI outputs is a skill that must be developed over time.

Setting up a GenAI management framework for your organization

General AI guidelines

The first task when setting out to use AI for everyday tasks is setting up general processes and rules around acceptable AI use for your organization. This should include both *do's* and *don'ts*: defining how to use AI safely, what kind of data can be fed to models, and specifying unacceptable uses of GenAI or uses that require prior approval. These organization-wide rules should be developed in partnership with IT and legal teams, and revisited with some frequency as model capabilities change. AI tools should also undergo a security and legal review to make sure their use of user-provided data complies with privacy and legal requirements for your organization. Most paid services for commercial models have options where they do not train on your data, but oftentimes these are “opt-out” rather than “opt-in”, i.e. you must explicitly say you do not want your data used for training.

Making the leap to custom + productionized GenAI

To move beyond out-of-the-box tools and models to political use cases, there is additional general setup required to ensure that models get the context they need. In general, organizations should seek to establish a “sandbox” for GenAI development (either in-house or with a vendor) that is flexible enough to serve as a starting point for many tasks your organization might want to solve or augment with AI, rather than building up significant infrastructure for one-off tooling.

- **Organization specific prompts:** Organizations should build out skeletons of what every person should know at the organization, including database and codebase documentation, style guides, general background knowledge, and any other information that would normally be passed formally or informally to new human employees. This is helpful to have in general and can then be used as model context. Task-specific instructions and context can then be layered on top of this org-wide context so that a model has both wide and narrow context to complete a task.
- **AI sandbox:** For tooling that requires custom pipelines or development, it may be helpful for organizations to designate a set of models and space to securely host tools. Scaling beyond simple chatbots will likely require API or programmatic access to model endpoints. You should decide which provider or providers to utilize, how to delegate access across your organization, and how to manage costs. Without a sandbox in place, every possible AI tool at your organization could face significant startup costs.
- **Fine-tuning LLMs:** Depending on use case, you may want to consider fine-tuning an existing model on a specific task when you have large volumes of domain-specific data and need consistent performance on repetitive workflows. Fine-tuning works best for tasks where you can provide many examples of correct inputs and outputs, such as standardizing document formatting, categorizing data, or generating content that follows specific organizational style guidelines. Rather than repeatedly providing extensive context in each prompt, fine-tuning bakes that knowledge directly into the model, making each individual request more cost-effective and reliable.
- **Structured output:** Several models and tools allow users to constrain the output of LLM calls into pre-defined schemas. Depending on the use case, this can help ensure that a given model only returns the specific output you’re looking for, and that this output is usable by any downstream systems.
- **Tests and success criteria:** As your organization develops tools and workflows, take time to define what a good example of the task looks like. This could be an example report, code snippet, etc. used as context for the model. Ideally organizations would also have “tests” for models to pass – this could be an actual pass/fail rate for tasks where judging success is black and white, or could involve something like a rubric. These tools can be used to judge the output of a model when monitoring model performance over time or evaluating whether to switch or upgrade models.

The ecosystem is evolving quickly - make sure you build a framework that can survive having the back-end model swapped out as newer, more capable models are released.

Use Cases

The following sections examine AI applications across three critical campaign functions where we've conducted systematic testing. Rather than broad theoretical assessments, these evaluations involved administering actual hiring assessments to frontier models and running AI tools through real campaign workflows to understand both capabilities and limitations.

Each use case follows a similar pattern: AI excels at routine, technical tasks with clear success criteria but requires significant human oversight for work involving judgment, strategy, or organizational context. The key is identifying where AI can meaningfully augment human work without introducing unacceptable risks or quality degradation. Model capabilities are moving very fast, so we recommend experimentation across the board to help find the shifting limits of this frontier.

Software engineering

For software engineering tasks, AI delivers its biggest wins when the work is both routine and easily verifiable. The productivity gains can be substantial: generating dbt YAML files that used to take hours of tedious manual work now happens in minutes. Similarly, AI excels at bridging skill gaps. Turning a backend script into a polished frontend interface no longer requires JavaScript expertise, and writing decent Terraform configurations is accessible even if you're not fluent in HashiCorp's syntax.

AI is particularly excellent at testing, both generating test cases and writing test scaffolding itself. Where test-writing was once tedious enough to skip, AI can rapidly generate both test structures and realistic fake data, leading to significantly better test coverage with less developer effort.

However, the limitations become apparent as complexity increases and verification becomes harder. While agentic "build, test, read logs, iterate" loops sometimes work, AI often gets stuck chasing rabbit holes or missing the forest for the trees. Much of software engineering, and especially data engineering, involves understanding system architecture, scoping problems correctly, and verifying outputs against business logic. This cognitive work still requires human oversight. More concerning are the security risks: AI suggestions can be dangerously naive, like recommending the removal of authentication layers to "fix" auth bugs. For infrastructure and security tasks, never run commands you don't fully understand.

Analysis and data science

To test AI capabilities in analytical roles, we administered our standard hiring assessments for data engineering, data science, and analyst positions to a number of frontier models. The results were revealing: AI excels at technical execution and theoretical knowledge but struggles with interpretation and judgment.

On the data engineering assessment, which focused on code review and query optimization, AI performed exceptionally well. It correctly identified performance issues, flagged problematic SQL patterns, and suggested meaningful improvements. This mirrors AI's broader strength in software engineering: when the task involves critiquing existing code against established best practices, AI delivers strong results.

The data science assessment told a more complex story. AI demonstrated solid theoretical knowledge and technical competency, handling statistical concepts well, communicating technical ideas clearly to non-technical audiences, and building reasonable models. However, cracks appeared in areas requiring nuanced judgment: it missed some weighted calculations and made errors in probability problems despite understanding the underlying concepts.

Most concerning was the analyst assessment, where the failure mode was particularly instructive. When asked to validate a machine learning model's performance, AI correctly plotted model scores against outcome data but then concluded the model was performing well despite the chart clearly showing no correlation between predictions and reality. AI created the right visualization but completely misinterpreted its own work.

This pattern extends to day-to-day analytical work. AI excels at exploratory data analysis and visualization tasks, quickly generating useful charts and summary statistics. It's particularly valuable for small verification scripts and data utilities, and it potentially unlocks entirely new frontiers of unstructured and qualitative data analysis However, three distinct limitations emerge for day-to-day analytical work.

First, there are practical tooling issues: AI sometimes fails to access uploaded files despite having Python capabilities, can't easily query organizational databases, and generally lacks the ergonomic integrations needed for smooth analytical workflows. These are solvable problems that better platforms and tools will likely address.

Second, analysis work requires a high level of detailed verification to ensure accuracy. Unlike debugging a piece of traditional software, where the main questions might be “does it run” or “does the frontend look like I want”, inaccurate analysis can often look deceptively compelling at first glance. Verifying analysis code involves not just ensuring accuracy of the outputs, but every step along the process. Did the model make the right assumptions when it dropped null entries, recategorized values, or aggregated data? This might involve a tedious, line-by-line review of generated SQL or Python, and the cost-benefit balance of “write it manually” vs. “generate with AI” may start to tip back in favor of the manual route.

Finally, and most fundamentally, AI consistently requires human oversight for any analysis requiring strategic judgment or business context. The model validation example illustrates this perfectly: AI can execute the technical steps of analysis but struggles to draw meaningful conclusions from the results. This isn't a tooling problem but a core limitation in how AI interprets data in context.

It's also worth noting that analytical AI use cases typically require sharing your organization's potentially sensitive data directly with AI models. Campaign databases, voter files, and internal polling data represent significant security risks if transmitted to third-party services. Open-source models that can be run locally may alleviate some of these concerns, but data security remains a critical consideration when evaluating AI for analytical workflows.

Hiring

The hiring process presents one of AI's most immediate challenges for campaign organizations. Our assessment experiments revealed a fundamental tension: AI performed well enough on technical evaluations to potentially reach second-round interviews, while simultaneously raising concerns about candidates masking gaps in understanding.

The Core Dilemma
Campaign organizations face two imperfect approaches to AI in hiring assessments:

Allow AI use but require disclosure. This approach acknowledges the reality that candidates will likely use AI regardless and attempts to evaluate AI collaboration skills alongside technical competency. However, it creates significant risks for overworked hiring managers. AI-assisted submissions often appear polished and comprehensive at first glance, potentially masking fundamental knowledge gaps. Given that effective AI use requires strong foundational understanding to validate outputs properly, candidates who rely too heavily on AI assistance may struggle once hired.

Prohibit AI usage entirely. This approach faces serious enforcement challenges, particularly for take-home assessments where monitoring is impossible. More problematically, it's somewhat hypocritical: campaigns will expect new hires to use AI effectively from day one to boost productivity. Restricting AI during hiring means missing candidates who are already skilled at AI collaboration - precisely the people campaigns want to hire.

Practical Recommendations
The enforcement challenges and evolving industry norms (Meta recently announced allowing AI in some coding interviews) suggest that adaptation rather than prohibition is the more sustainable approach. For campaign organizations, this means:

Shift toward live assessments, or at the very least, live walkthroughs of assessments. Rather than take-home tests that are impossible to monitor, conduct more live coding sessions, analysis walkthroughs, or problem-solving conversations. This allows you to observe how candidates use AI tools, whether they understand the output, and how they verify results. We've heard from some folks who are successfully combining take-home and live assessments, by assigning "open book"/"open AI" take-home assessments, and following them up with a walkthrough interview. The take-home still allows the candidate to work through meatier problems with reduced time-pressure, but the interviewer still gets an opportunity to ask questions about the produced work and ensure the candidate understands the code that's been either written or generated.

Test AI collaboration explicitly. Given the efficiency gains at stake, hiring managers should seek out candidates with experience using AI and assess candidates' ability to use it well. Ask them to explain their AI-assisted process, critique AI-generated outputs, or identify potential errors in AI-produced work.

Focus on fundamentals that AI can't replace. Emphasize assessment components that require judgment, strategic thinking, and contextual understanding. For analysts, this means testing their ability to draw meaningful conclusions from data rather than just generate charts. For engineers, focus on architectural decisions and debugging skills rather than pure code generation.

The reality is that campaign hiring managers, already stretched thin, will face an influx of polished AI-assisted applications. The organizations that adapt their hiring processes to explicitly evaluate AI collaboration skills while maintaining standards for fundamental competency will be better positioned to build effective teams in an AI-augmented environment.

Acknowledgements

Thank you to Switchboard and Zinc Collective for their support for this project.

Thank you also to the following readers for their generous and thoughtful feedback: Valerie Bradley, Anna Carmichael, Kevin Collins, Christina Coloroso, Johannes Fischer, Arpit Garg, Becca Siegel, Jack Solano, John Spaw, Matthew Stafford, Benn Stancil, Maeve Ward.