

AI benefits abundance and public understanding

Evidence and public communications from 10/05/23 through 10/05/26

- The shortcomings in public explanations of AI involve translation, delivery, and distribution.
- Descriptions of technological capability do not by themselves establish which ordinary problem it solves, what a person must do to obtain the benefit, and who receives the resulting gains.

Everyday need	Useful assistance	What would count as a benefit
Understand a difficult letter	Explain the supplied text, extract dates, list unanswered questions	Correct comprehension and fewer missed deadlines
Write something important	Organize notes, draft, revise tone, check completeness	A better message with less effort, retaining the person’s intent
Learn a topic	Explain alternatives, give practice, diagnose a mistaken step	Knowledge retained when the assistant is removed

Everyday need	Useful assistance	What would count as a benefit
Speak across languages	Translate, explain idioms, rehearse a conversation	More successful communication, with checking for consequential details
Manage family logistics	Turn emails and notices into a proposed calendar and checklist	Fewer omissions and less planning time after review
Compare purchases	Compare total costs, constraints, evidence, and return conditions	A suitable purchase and net savings after fees and checking
Prepare for an appointment	Organize records and questions for a professional	A clearer, more productive consultation
Apply for work	Tailor truthful experience to a role and practice interviews	Stronger preparation; offers require separate outcome evidence
Run a small organization	Draft correspondence, reconcile supplied records, prepare materials	Less administration and more time serving clients or members
Create for enjoyment	Develop stories, invitations, images, music concepts, or hobby plans	Enjoyment and expressive capability, not necessarily income
Access visual information	Describe images and inaccessible text through a suitable service	Greater independence with a route to human assistance

Everyday need	Useful assistance	What would count as a benefit
Pursue a continuing goal	Monitor relevant updates and prepare the next step	Useful follow-through with fewer interruptions and errors

These benefits have several distinct forms.

AI may save time on something a person already does; improve quality at similar effort; let the person do something previously unaffordable; reduce the difficulty of getting started; or increase independence.

Treating all of these as “productivity” hides much of their appeal. A person who can finally understand a school notice or create a family history may gain value even if measured output and income do not change.

There are also indirect benefits. A person may receive better customer support, a clearer public-service notice, or a faster response from a small business because the organization uses AI. For those benefits, the relevant evidence is the recipient’s experience, not the organization’s volume of generated text.

Scientific discovery and improved physical production are another pathway, but they depend on research systems, experiments, deployment, and material capacity; they cannot simply be credited to a consumer chatbot.

How to read the evidence

Evidence comes in different strengths. A randomized trial can estimate an effect within its setting. A workplace rollout may supply valuable real-world evidence without full randomization.

Usage logs reveal what people do, but not what would have happened without AI. Surveys measure perceptions or reported behavior. Product demonstrations establish that a selected workflow can be shown; advertisements establish what a company wants audiences to believe. Forecasts describe conditional futures. None should silently substitute for another.

The practical question is whether the whole interaction helps.

Compare the result with the best realistic alternative, including search, a template, conventional software, a colleague, or professional help. Include setup, prompting, waiting, checking, correction, subscription charges, and the consequences of mistakes. An impressive draft that takes longer to verify than to write can have negative value for that task.

The strongest findings are useful precisely because their boundaries are visible.

Evidence and setting	Result	What the result does not establish
Customer support rollout, 5,172 agents; published 2025	15% more issues resolved per hour; about 30% for less experienced or less skilled workers	A universal productivity increase across jobs or industries [5]
Professional writing experiment, 453 professionals; July 2023	40% less task time and 18% higher assessed quality	Complete-job performance or long-term income gains; this foundational study predates the review window [6]
Consulting experiment, 758 BCG consultants; published 2026	On suitable tasks, 12.2% more completed with 25.1% less completion time; on a difficult task, a 19-percentage-point lower correctness rate	Reliable judgment whenever an answer sounds coherent [7]
P&G product-development experiment; published 2026	Individuals with AI matched teams without AI on assessed solution quality	Replacement of sustained teamwork or improved ability to select the best idea [8]
METR developer experiment, 16 developers and 246 tasks; 2025	With early-2025 tools, completion took 19% longer in familiar repositories	A current general verdict on coding tools; 2026 follow-up estimates were inconclusive and affected by selection problems [9, 10]

Evidence and setting	Result	What the result does not establish
Harvard physics tutoring crossover, 194 eligible students; published 2025	A specially designed tutor improved immediate learning on two lessons	General chatbot superiority over teachers, or demonstrated long-term retention [11]
Turkish mathematics experiment, nearly 1,000 students; published 2025	ChatGPT-style GPT-4 assistance improved practice but produced 17% worse later unassisted test scores	That all tutoring is harmful; guided tutor design largely removed the deficit but did not establish an unassisted-test gain [12]
Nigerian public-school program; 2025 working paper	Six weeks of AI tutoring plus teacher support improved the combined assessment by 0.31 standard deviations	A chatbot-only effect or literally two years of observed learning [13]
Randomized short-story study; published 2024	AI ideas improved individual assessments, especially for participants with lower initial creativity scores, while making stories more similar	Universal creative improvement or long-term cultural enrichment [14]

Evidence and setting	Result	What the result does not establish
UK medical-scenario experiment, 1,298 participants; published 2026	LLM access did not improve selection of the appropriate level of care	That a model's medical knowledge score translates into better public decisions [16]

Mental-health assistance illustrates why product specificity matters. A 2025 trial randomized 210 adults to a purpose-built research chatbot, Therabot, or a waitlist and found symptom improvements. It used an expert-trained system and oversight, with limited follow-up. It did not compare the tool directly with human therapy. The frequently publicized percentage symptom reductions are within-treatment changes, not the randomized between-group effect. General chatbot companionship should not inherit the trial's clinical claims. [15]

Usage research adds breadth to the experiments. A company-linked NBER study analyzed 1.5 million ChatGPT conversations and found roughly 70% of consumer use was nonwork; practical guidance, information seeking, and writing dominated. This is evidence that ordinary-life assistance matters to users, not proof that every interaction benefited them. Because it covers one platform and uses automated classifications, it should not be treated as a census of all AI use. [17]

A Danish study linked surveys of about 25,000 workers to administrative data. Users reported savings averaging 2.8% of working hours, while early effects on earnings and hours were small. A separate 2026 Stanford working paper estimated substantial US consumer value from willingness to accept

compensation for losing chatbot access. Its mean valuations were much higher than its medians, and it relied on stated choices and extrapolation. Together these studies show why saved effort, subjective value, wages, and GDP must remain separate measures. [18, 19]

Accessibility makes the capability benefit unusually concrete. Be My AI can describe photos and image-based text and offers a route to a sighted volunteer. The service's documentation also explicitly excludes reliance on it for medicine labels or dosage. The useful claim is improved access to some visual information, with support and boundaries. The existence of this workflow is stronger evidence than a generic promise to empower everyone, though it is not by itself a population-level outcome evaluation. [20]

The costs and failure modes also differ by use.

- For factual questions, invented details and confident omissions can make verification expensive.
- For learning, immediate success can conceal dependence.
- For creative work, easy generation can increase similarity and verification demands.
- In connected workflows, mistaken actions and excessive access can matter more than a wrong sentence.
- At population scale, unequal access, labor displacement, misinformation, and infrastructure burdens can coexist with individual gains. Their magnitude and distribution require specific evidence; listing a possibility does not establish how often it occurs.

Several workplace studies use company partners or company-linked authors; the settings are unusually prepared and the tasks often bounded.

Educational interventions can require substantial teacher preparation. These facts do not invalidate the results. They identify the investments that a real deployment may need to reproduce them. Newer models may improve results, but that improvement requires fresh outcome measurement rather than being assumed from a newer product name.

What company messages communicate

The claim that AI companies never explain benefits or define abundance is too strong. Their communications include detailed positive visions, concrete consumer demonstrations, and increasingly explicit policy proposals. The more defensible criticism concerns the completeness and reach of the explanation, and the gap between illustration and verified delivery.

An early statement within the review period is Marc Andreessen's October 2023 manifesto, which defines abundance through falling prices. It makes the market-transmission argument explicit, but it is investor advocacy rather than evidence that the full process will occur. Its confidence in the social effects should be assessed separately from the usefulness of the price criterion. [57]

Sam Altman's 2024 essay *The Intelligence Age* describes personalized assistance and tutoring alongside far-reaching scientific and economic possibilities. Dario Amodei's October 2024 *Machines of Loving Grace* offers a much more extensive account of possible gains in health, biology, economic development, governance, work, and meaning. These are substantive visions. Their most dramatic outcomes depend on future capabilities and complementary changes, so they should be presented as scenarios rather than delivery commitments. [21, 22]

Altman's June 2025 *The Gentle Singularity* connects falling intelligence costs and technological progress to abundance while acknowledging distribution and institutional problems. His February 2025 *Three Observations* also raises the issue of uneven gains and access. Meta's 2025 personal-superintelligence statement emphasizes helping individuals pursue their interests and lives. These sources reveal different emphases, but none supplies a demonstrated household entitlement to cheaper necessities or more disposable income by a fixed date. [23, 24, 25]

OpenAI CFO Sarah Friar provides a particularly important counterexample in July 2026. Her definition of abundance is intelligence that becomes more capable, affordable, and valuable. She explicitly identifies the cost of a successful outcome, including retries and oversight, as the relevant economic measure. That is a useful operational definition. It still leaves a separate public question: how much of the resulting value reaches people as lower prices, greater access, better pay, or more free time? [26]

Google DeepMind's September 2026 economic-policy essay addresses that second question directly. It considers phased income support, training, and broader capital ownership, linked to observed disruption. Its policy ratings partly use simulated economist personas, not a vote by 51 independent human experts or a trial of the policies. OpenAI has also published industrial-policy proposals, including a public wealth fund, a tax base less dependent on labor income, wider access to AI, and shorter-workweek pilots without reduced pay.

These are meaningful efforts to address distribution; they remain proposals and research programs, not proof that society has adopted effective arrangements. OpenAI's August announcement of 14 policy-research grants

is observable follow-through, with results expected in 2027 rather than established impacts now. [27, 28, 64]

Advertisements reveal progress in making the use case visible. OpenAI's 2025 Super Bowl campaign placed AI within a long history of transformative tools. Google's 2025 Dream Job campaign showed a father practicing for an interview, making the personal objective and interaction easier to understand. OpenAI's later everyday-life campaign used cooking, travel, and exercise. Anthropic's Keep thinking campaign emphasized human reasoning, and its 2026 Space to think message made an explicit commitment to keeping ads out of Claude conversations. [29, 30, 31, 49, 50]

These examples serve different purposes. A historical montage can establish importance; a familiar task can explain use; an incentive commitment can address trust. None alone proves a lasting benefit. An everyday example also needs to show why AI is preferable to existing search or a simple template. The useful differentiator might be adapting to several constraints, responding to follow-up questions, or completing the next step. Where there is no advantage, conventional tools may be the better choice.

Anthropic's July 2026 Hard questions campaign explicitly addresses public concerns about jobs, children, agency, and other consequences. That acknowledges that reassurance must include difficult issues. Whether such advertising changes understanding or trust requires audience research; its existence does not establish success. Meta's September small-business Muse announcement similarly uses a recognizable trade and administrative burden, but selected testimonials cannot substitute for representative outcome evidence. [32, 51]

The reviewed communications leave several questions unevenly answered. Does a technical metric predict a useful result? Is a demonstration representative of normal performance? Does access extend across language, disability, time, money, and connectivity differences? Does a promised date refer to a software release or a working service? Does increased output give the person more time or higher expectations? Who receives a projected economic gain?

These are diagnostic questions for evaluating a message. This review does not establish the prevalence of each omission across all industry advertising.

What Public Services Alliance contributes

Public Services Alliance is especially relevant because its recent posts directly examine these gaps. Its August 2026 post about Altman and public acceptance reproduces reporting that he acknowledged the difficulty of changing habits and communicating benefits. That is secondhand reporting of an interview, so it should not be treated as an independently checked transcript. [33]

PSA's September post A finished tool is not proof it works makes a useful distinction between building something quickly and demonstrating better learning or care. Its June post on AI fluency emphasizes supported practice, feedback, checking, and escalation. Its January discussion of hidden assumptions asks whether costs, authority, dignity, and incentives have been evaluated. Several posts use or reproduce AI-assisted analysis. They contribute a public-service framework and leads to sources; they are not independent trials validating AI outcomes. [34, 35, 36]

The July post *AI Access Is More Than Availability* supplies a practical test for “AI for everyone”: can people benefit given their devices, connectivity, skills, privacy needs, and access to help or recourse? The same month’s data-center discussions connect public-service promises to local electricity and water burdens. A community can reasonably ask for specific service commitments and a fair allocation of infrastructure costs. This makes the issue partly one of delivery and accountability, rather than publicity alone. [52, 53, 54]

Earlier PSA posts show the same tension developing over time. Its December 2024 *Second Machine Age* discussion combines enthusiasm for cognitive assistance with uncertainty about the sources behind an AI-generated analysis. A May 2025 *LearnLM* post explains educational uses, but supplies no independent outcome trial. Its August discussion of AI interfaces emphasizes continuity and memory; November coverage discusses literacy; and a December repost of David Shapiro argues that concentrated infrastructure and ownership may restrict the public promise. These are attributed interpretations, not empirical confirmation of those claims. [58, 59, 60, 61, 62]

The retrieved PSA material is selected coverage from 2024–2026. Relevant 2023 archive pages could not be reliably retrieved, so this report does not claim a complete audit of every PSA post across the three-year period. Public education should therefore do more than list prompts. People need guided chances to complete a task they care about, compare the result with a reliable source, and decide when the tool is inappropriate. Libraries, adult education, schools, disability organizations, and frontline public services could provide this support. That is a proposed implementation approach; its local design and effectiveness would need evaluation.

A precise definition of abundance

For this report, AI-created abundance means a sustained increase in people’s effective access to useful knowledge, assistance, goods, and services at acceptable quality, requiring less money, time, or effort, with gains reaching ordinary households and preserving meaningful choice.

This is a proposed public-interest definition. It is intentionally stronger than a definition based only on cheaper model access. It allows improvements in quality and access as well as lower prices. It also requires examination of distribution: an average gain can coexist with some people becoming substantially worse off.

Five forms of abundance should be named separately.

Form	What becomes more available	Evidence or condition to look for
Cognitive assistance	Explanations, drafts, translation, analysis, coding, practice	Cost and reliability of a useful answer or artifact
Personal capability	Ability to undertake projects once requiring specialist help	New tasks successfully completed by ordinary users
Service capacity	Education, administration, customer support, care support	Better outcomes, shorter waits, and affordable access
Material provision	Medicines, energy, housing, food, transport	Real production, deployment, and affordability of final goods

Form	What becomes more available	Evidence or condition to look for
Shared prosperity	Purchasing power, security, discretionary time, agency	Gains across income groups, regions, and abilities

The progression is conditional. A model must be capable of the relevant task. The product must perform it reliably. People and organizations must adopt it and change their processes. Its total cost must fall or quality must rise. Supply must expand where needed. Benefits must reach users through lower prices, better service, higher earnings, public provision, or other arrangements. Only then is a broad claim about household welfare established.

A simple numerical example shows why small task gains cannot be scaled casually.

Suppose administration accounts for 10% of a service's total cost and AI halves that administrative cost. The direct saving is 5% of total cost before integration and checking. If a competitive or regulated provider passes through all of that saving, the price could fall by 5%, all else equal. A 50% task improvement is not a 50% reduction in the final bill. This example is arithmetic, not an empirical estimate of any particular service.

Time also needs careful accounting. In an illustrative household workflow, reducing a chore from 60 minutes to 20 minutes of work plus 10 minutes of checking releases 30 minutes. If access problems and corrections consume another 40 minutes, it instead costs 10 minutes. Even when time is saved, it becomes leisure only if the person can retain it. At work, an employer may use the capacity to increase output. In a household, avoided mental effort or fewer missed obligations can matter independently of minutes.

Price abundance and income security must be evaluated together. If a household's relevant expenses fall by 10% but its income falls by 30%, cheaper goods do not make that household better off. Nor does more content automatically make information more useful: attention, verification, trust, and human contact can become the scarce resources. Original locations, scarce natural resources, social status, and other positional goods do not become unlimited because language processing becomes cheaper.

AI can help with planning a house, researching a medicine, or managing an electricity grid. It cannot by itself supply the land, clinical validation, construction crews, manufacturing capacity, distribution systems, legal authority, or human agreement those activities require. Conversely, better models can help relieve some of these bottlenecks. **The appropriate claim is conditional acceleration, with evidence at each stage, rather than automatic removal of scarcity.**

What the economy currently shows

The macroeconomic evidence does not justify saying that nothing is happening. It also does not establish economy-wide abundance. An April 2026 BEA working paper finds an association between AI intensity and lower purchaser prices, partly connected with reduced input-cost contributions. Its authors explicitly place causal identification outside the scope of that early analysis. This is a promising signal, not proof that AI caused a general fall in household living costs. [45]

A July 2026 Federal Reserve research note distinguishes rapid capability improvement and investment from the still-limited aggregate evidence of broad transformation. It explains how integration costs, sector differences, shallow use, and measurement difficulties can interrupt the path from task gains to total productivity. The ILO's May 2026 research brief likewise emphasizes diffusion, skills, organizational change, competition, and worker participation. Those accounts support patience about aggregate measurement while requiring restraint about claims already made. [46, 47]

Infrastructure has a real cost. The IEA's 2026 update projects global data-center electricity consumption rising from about 485 terawatt-hours in 2025 to 950 in 2030. That includes all data centers, not only AI. It also finds rapidly improving task efficiency, alongside growth in use and more energy-intensive agentic and media workloads. The report treats effects on electricity prices as dependent on local capacity, investment, flexibility, and cost allocation. It supports neither an assumption of negligible costs nor a claim that every AI interaction imposes the same burden. [48]

The policy routes to shared gains deserve separate evaluation: competition can help transmit lower costs into prices; public procurement can purchase

useful services; accessible education can broaden effective use; labor institutions can influence pay and hours; income support can protect people during disruption; and public or broad-based ownership can share capital returns. Each route has implementation costs and trade-offs. Some combination may be needed, and appropriate choices depend on the observed pattern of change. None follows automatically from scaling a model.

A credible public abundance scorecard would report the following alongside adoption and investment.

Dimension	Metric worth reporting	Comparison needed
Affordability	Cost of a quality-adjusted useful service and cost of essentials relative to income	Prior cost, conventional alternative, and fees
Time	Time saved after setup, review, correction, and interruptions	The whole workflow, not generation speed
Quality	Correct completions, learning retained, recipient outcomes	Independent assessment and an appropriate control
Access	Successful use across age, language, disability, income, and region	Gaps between groups and rates of exclusion
Work	Earnings, job quality, hours, employment transitions	Distribution over time, not an average task score

Dimension	Metric worth reporting	Comparison needed
Agency	Ability to refuse, correct, export, appeal, and obtain human help	Actual usability of these choices
Community effects	Local infrastructure costs, resource use, and public benefits	Who pays, who benefits, and enforceable commitments

There is no defensible universal date at which all five forms of abundance arrive.

Cognitive assistance and selected capability gains are already observable. Broader service benefits require implementation and measurement. Material and shared prosperity depend on longer and less predictable transitions. Dates should attach to particular deliverables and access commitments, not to an undefined civilizational state.

Meta Muse as a case study

Meta launched Muse on September 8, 2026 as a personal AI agent with its own cloud computer and browser. It can continue work after the app closes and use connected services for tasks such as shopping, forms, and travel preparation.

It is distinct from the Muse Spark model family and from the earlier Meta AI app. It launched in the United States and subsequently expanded, with regional availability still limited. A free usage allowance and paid higher-

usage plans lower the barrier to trying it, but free entry does not imply unlimited work. These are documented product features and access conditions, not representative success rates. [32, 37, 55]

Muse's design account identifies the discovery problem unusually clearly: early testers did not know where to begin. Meta responded with personalized onboarding suggestions and an Ideas view. Its broader design combines persistent context, concurrent work, editable memory, visible goals and activity, structured approvals, and outputs suited to the task. A customizable avatar gives the interaction a familiar identity. These features can make an abstract capability easier to grasp and supervise, although their effects on adoption have not been isolated experimentally. [38]

The small-business expansion adds concrete workflows such as reviewing expenses, preparing marketing materials, and tracking unanswered correspondence. That framing speaks to the work an owner wants to get back to. An independent TechRadar reviewer also described useful retrieval and draft-response assistance while finding the required account access uncomfortable. The latter is a single hands-on experience, valuable as an illustration rather than an estimate of reliability. [32, 40]

Early popularity is supported by installation estimates. On September 30, 9to5Mac reported Sensor Tower's projection that Muse had crossed five million downloads and had spent twelve consecutive days at the top of the US App Store.

The same report documents a large advertising push, including up to half of Meta's daily house-ad impressions over a recent two-week period. These are attributed third-party estimates. They establish substantial trial and attention; they do not establish retention, representative satisfaction, or net

time saved. Cross-app launch comparisons also differ in platform access, geography, and historical conditions. [39]

Several contributors to uptake are therefore plausible: recognizable tasks, less pressure to invent a good prompt, persistent assistance, familiar presentation, free entry, and Meta's existing reach. Advertising may drive trial independently of usefulness. Useful results may drive repeat use independently of advertising. Without acquisition attribution and retention data, the relative contribution of each mechanism remains unknown. As of this report's cutoff, Muse is less than a month old.

Its most important limitations concern execution and trust. Connected accounts provide the context needed for help but also increase the consequences of errors or inappropriate disclosure. Website restrictions and counterparties' willingness to accept agent actions can limit execution. Purchases, messages, and applications require clear authority. Meta describes isolated execution, protected credentials, separate permission enforcement, and review of consequential actions; these are safety design claims, not guarantees against failure. [41]

A revealing discrepancy appears between Meta's simple launch presentation and its getting-started guidance. The latter asks users to supply context, set boundaries, and monitor unexpected or inaccurate behavior. A useful assistant can reduce work while still requiring judgment. Communications should make that remaining responsibility legible before a person depends on it. [56]

How OpenAI and Anthropic can compete

Neither company is starting from zero. **OpenAI launched dots on September 29, 2026**, describing persistent cloud agents, memory, connected applications, and activity and permission controls. Initial access centers on Pro and Business Premium, with enterprise beta access, so it is not a universal free rollout.

Anthropic announced cloud Cowork on web and mobile in July; its current guidance also describes merging chat and Cowork, initially for Pro and Max. Cloud and local-computer tasks have different requirements. These products already overlap with major parts of Muse's approach. [42, 43, 44]

The strategic opportunity is to make their capabilities easier for ordinary users to discover, obtain, and trust, and to demonstrate better outcomes. The following are recommendations inferred from the evidence, not findings that a particular company has already validated.

Product choice	What to reproduce or improve	Evidence of success to require
First experience	Offer a small set of relevant tasks and a guided first success	Completion and repeat use among inexperienced users

Product choice	What to reproduce or improve	Evidence of success to require
Setup	Start with a supplied document or public information before requesting broad account access	Less abandonment and a useful result before extensive setup
Continuity	Preserve user-approved goals and context; make memory inspectable and correctable	Fewer repeated instructions and fewer incorrect assumptions
Background work	Continue authorized tasks and interrupt only for meaningful developments	Net time saved and low notification burden
Results	Deliver an itinerary, calendar, comparison, file, or verified action receipt	Correct and usable outputs, rather than persuasive completion claims
Control	Explain access and consequential actions through clear interface controls	Users understand, can revoke, and can recover from decisions
Access	Affordable entry, accessible interfaces, language support, community instruction	Benefits beyond existing professional and technical audiences

Product choice	What to reproduce or improve	Evidence of success to require
Evaluation	Publish success, failure, checking time, cost, and retention by task and user group	Reproducible benefit over realistic alternatives

For OpenAI, the clearest opportunity is to package its broad tool and agent capabilities around household and community needs, with simpler entry than a menu of model and product names. A user should be able to begin with “help me keep up with these school notices” and understand what will be produced, what access is needed, and what remains for them to approve. Whether that workflow belongs in ChatGPT, dots, or another surface is a product decision; the person should not need to understand the internal architecture first.

For Anthropic, the opportunity is to extend its emphasis on deliberate work and user trust into familiar personal and small-organization tasks. Combining conversation and action can remove a confusing mode choice. Its ad-free conversational commitment is a potentially meaningful distinction when people ask whether recommendations serve their interests. That commitment alone does not establish privacy, accuracy, or complete alignment; each needs its own evidence. [50]

For both companies, copying a mascot would capture only a small part of the lesson. A friendly identity may reduce intimidation, but the durable advantage would be lower effort to obtain a reliable result. Neither can simply duplicate Meta’s existing consumer distribution. Partnerships with

trusted organizations, useful shareable outputs, and clear pricing offer alternative ways to reach people, subject to testing.

A convincing comparison should:

- give all three products the same ordinary tasks, the same account permissions, and comparable budgets.
- Recruit new users as well as experienced ones.
- Evaluate correct completion, user effort, total elapsed time, money spent, failures, unwanted actions, and understanding of permissions.
- Follow users long enough to separate novelty from sustained value.
- Include age, language, disability, and connectivity differences.
- Report individual results and variation, not only a single winner.

What an honest benefit explanation should contain

Commercial incentives belong in the explanation too. OpenAI's October 5, 2026 advertising announcement argues that ads support broad access and says they remain separate from, and do not influence, ChatGPT answers. That is a company commitment which users and independent researchers can evaluate. Providers should distinguish affordability supported by advertising from evidence that every recommendation serves the user, and should make sponsorship and data practices understandable. [63]

A strong public explanation starts with a burden a person recognizes. It shows the actual input, the system's contribution, the usable result, and the remaining human responsibility. It states which feature is available, at what price, in which circumstances, and with what limits. It distinguishes a

demonstration from a measured claim and a future possibility. Companies should measure whether viewers can explain the benefit accurately and complete the task, not merely whether they remember the brand.

For example, a campaign about family administration could show several school notices becoming a proposed calendar, with dates linked to the original notices and uncertain details flagged. It should show the parent reviewing the result before sharing it. The public claim should be that the tool helps organize the material, unless a study also supports a specific average time saving or reduction in missed events. A claim of fewer missed deadlines requires follow-up evidence.

A learning campaign should show a student working through a misconception and later answering a new problem without AI. A small-business campaign should show an owner inspecting a reconciled record and the amount of correction required. An accessibility campaign should show both independent use and the human fallback. These are proposed demonstration formats informed by the evidence; they are not reports of tests already conducted.

The full public promise should have four parts:

- a useful capability now,
- evidence of its benefit in specified conditions,
- a candid account of its costs and remaining responsibilities,
- and a credible route for broader access.

This would give people a basis for choosing, questioning, or declining the technology. It would also make future claims about abundance accountable to changes in their lives.

Research gaps and next tests

The most consequential unknowns are the net value of consumer agents across representative users, durable educational effects, the conversion of task savings into earnings or leisure, and the distribution of gains and costs. Muse in particular needs follow-up evidence beyond launch downloads. Older causal studies should be replicated with current products while preserving realistic tasks and including checking costs. Evidence should also include people who stop using a tool or cannot use it successfully.

A useful next research program would run three parallel studies. First, compare communication approaches by measuring comprehension and successful first use. Second, compare full product workflows over several weeks, recording errors and time rather than relying only on self-reports. Third, track household and community outcomes such as costs, waiting times, income, independence, and infrastructure burdens. None should assume that adoption itself is the desired outcome.

The strongest public case for AI is already available in narrower, testable terms: assistance can make certain activities easier, expand what people can do, and improve selected services. The larger claim becomes credible when those benefits are dependable, affordable, broadly accessible, and visible in household outcomes. Public trust will depend on that record as much as on the quality of the explanation.

These are related failures, but they require different remedies. Better advertising can explain a task. Better products can complete it. Broad prosperity also requires institutions, investment, competition, and decisions about income and ownership.

The evidence supports meaningful benefits already:

- help with writing and understanding information,
- some forms of learning, accessibility, selected work tasks,
- and the administration of everyday life.

It does not establish dependable improvement on every task, universal household savings, or a dated arrival of general abundance. Meta Muse is a useful recent case because it packages AI around recognizable chores and continuing assistance. Its rapid uptake occurred alongside extensive distribution and promotion; their contribution cannot yet be separated from product appeal. Its short history cannot establish lasting public benefit.

This report tests the two proposed failures rather than assuming them. It reviews published research, working papers, surveys, official product descriptions and advertisements, economic analysis, Public Services Alliance posts, and public education material. It is a broad, source-grounded synthesis rather than an exhaustive systematic review of every publication.

Most causal evidence concerns specific tasks and older model versions; many studies concern the United States or digitally connected populations. Source dates, study status, and important limits are identified locally. Recommendations and the proposed definition of abundance are this report's analysis, not measured findings.

What the public evidence establishes

The United States provides a clear illustration of adoption without broad reassurance. Pew's February 2026 survey of 5,119 adults, published in June, found that 49% reported using chatbots and 24% used them daily. Yet 40% expected AI's twenty-year effect on society to be negative, compared with 16% expecting a positive effect. Thirty percent of all adults said chatbots helped their productivity, versus 5% saying they hurt it. These answers are compatible: a useful personal tool can coexist with concern about jobs, privacy, or collective outcomes. The survey measures attitudes and reported behavior, not causal benefits or advertising effectiveness. [1]

Among nonusers, 60% named lack of interest as a major reason, 54% privacy concerns, 45% accuracy concerns, and 29% not knowing how to use chatbots. Multiple answers were allowed. Teaching people how to prompt addresses only one of these barriers. An explanation of benefits must earn interest, while product design and business practices must address the reasons for distrust. It would be unjustified to interpret all resistance as ignorance. [2]

The general population is not a single audience. AARP's January 2026 synthesis reports strong interest among adults over 50 in practical applications including translation, learning, and independence, with less interest in emotional companionship. Some of the applications in that research concern AI broadly, not LLMs specifically. Pew's international research also shows substantial differences among countries, so US skepticism should not be presented as a universal global verdict. [3, 4]

The practical diagnosis is a gap between making a tool available and making its benefits discoverable, verifiable, and equitably accessible. That judgment

follows from the combined evidence and communication examples; the reviewed material does not isolate poor messaging as the causal explanation for public skepticism.

What an LLM actually adds

An LLM is a system trained to process and generate language and, in many current products, images, audio, and other information.

- A chatbot provides a conversational interface.
- An agent adds tools, access to other services, memory, and the ability to continue a sequence of actions.
- A successful product also needs usable controls, reliable integrations, and a way to recover when something fails.
-

These layers matter because a strong answer in a chat window is different from a correctly completed application, purchase, or appointment.

For an ordinary person, **the useful change is often a reduction in the effort between a need and a usable result.**

- The person can ask follow-up questions,
- specify constraints, request a simpler explanation,
- rehearse a conversation,
- or turn a rough idea into a draft.
- The output can be tailored to the person's circumstances without commissioning a new piece of software or hiring help for every small task.

Quality varies, and fluent language is not a guarantee of correctness.

Sources and evidence status

Numbered references in the text link directly to the relevant source. Entries identify evidence type so that a product claim, opinion, survey, experiment, and forecast are not treated as interchangeable. Dates describe publication or the stated access date; experiments may have used earlier model versions. Some publisher pages required author-hosted copies or primary-source indexed text for access. PSA coverage is selective rather than an exhaustive archive audit.

[1] Pew Research Center. [Americans and AI 2026](#). June 17 2026. Representative US survey.

[2] Pew Research Center. [Why don't people use chatbots](#). June 17 2026. Nonuser survey findings.

[3] AARP. [Navigating the World of AI](#). January 2026. Survey synthesis covering AI broadly.

[4] Pew Research Center. [How People Around the World View AI](#). October 15 2025. International public opinion survey.

[5] Brynjolfsson Li and Raymond. [Generative AI at Work](#). May 2025. Published in Quarterly Journal of Economics.

[6] Noy and Zhang. [Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence](#). July 14 2023. Published in Science and outside main review window.

[7] Dell'Acqua and colleagues. [Navigating the Jagged Technological Frontier](#). March 11 2026. Published in Organization Science.

- [8] Dell’Acqua and colleagues. [The Cybernetic Teammate](#). June 12 2026. Published in Organization Science.
- [9] METR researchers. [Measuring the Impact of Early 2025 AI on Experienced Open Source Developer Productivity](#). July 2025. Randomized study and preprint.
- [10] METR. [Developer Productivity Study Update](#). February 24 2026. Primary follow-up with methodological limitations.
- [11] Kestin and colleagues. [AI Tutoring Outperforms In Class Active Learning](#). June 3 2025. Published in Scientific Reports.
- [12] Bastani and colleagues. [Generative AI without Guardrails Can Harm Learning](#). June 2025. Published in PNAS.
- [13] De Simone and colleagues. [From Chalkboards to Chatbots](#). May 2025. World Bank working paper on Nigeria.
- [14] Doshi and Hauser. [Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content](#). July 12 2024. Published in Science Advances.
- [15] Heinz and colleagues. [Randomized Trial of a Generative AI Chatbot for Mental Health Treatment](#). March 27 2025. Published in NEJM AI.
- [16] Bean and colleagues. [Reliability of LLMs as Medical Assistants for the General Public](#). February 9 2026. Published in Nature Medicine.
- [17] Chatterji and colleagues. [How People Use ChatGPT](#). September 2025. NBER working paper with company-linked authors.
- [18] Humlum and Vestergaard. [Large Language Models Small Labor Market Effects](#). 2025. Working paper linking Danish surveys and administrative data.
- [19] Brynjolfsson and colleagues. [What Is Generative AI Worth](#). April 13 2026. Stanford working paper using stated valuations.
- [20] Be My Eyes. [Using Be My AI](#). Updated July 1 2026. Official product guidance and limitations.
- [21] Sam Altman. [The Intelligence Age](#). September 23 2024. Executive vision essay.

- [22] Dario Amodei. [Machines of Loving Grace](#). October 2024. Executive scenario essay.
- [23] Sam Altman. [The Gentle Singularity](#). June 10 2025. Executive forecast essay.
- [24] Sam Altman. [Three Observations](#). February 2025. Executive economic and capability argument.
- [25] Mark Zuckerberg. [Personal Superintelligence](#). July 30 2025. Meta public vision.
- [26] Sarah Friar for OpenAI. [Building Abundant Intelligence](#). July 31 2026. Company definition and economic claims.
- [27] Julian Jacobs and Alex Imas. [Economic Policy for AGI](#). September 16 2026. DeepMind policy analysis partly using simulated raters.
- [28] OpenAI. [Industrial Policy for the Intelligence Age](#). April 2026. Policy proposals.
- [29] OpenAI Forum. [OpenAI Super Bowl Ad Introducing the Intelligence Age](#). February 12 2025. Company discussion of advertising.
- [30] Google. [Dream Job and Party Blitz](#). February 6 2025. Official advertising explanation.
- [31] Marketing Interactive. [OpenAI Shows How ChatGPT Fits into Everyday Life](#). September 2025. Campaign reporting.
- [32] Meta. [The Future Is for Everyone Muse for Small Business](#). September 29 2026. Official product announcement and testimonials.
- [33] Public Services Alliance. [Altman AI Tech Industry Has Not Yet Managed to Capture US Hearts and Minds](#). August 25 2026. Reposted reporting about an interview.
- [34] Public Services Alliance. [A Finished Tool Is Not Proof It Works](#). September 23 2026. Public-service commentary.
- [35] Public Services Alliance. [Training in AI Fluency](#). June 8 2026. Practical training commentary.

- [36] Public Services Alliance. [Hidden Assumptions and Meaning Gaps](#). January 20 2026. AI-assisted evaluation framework.
- [37] Meta. [Introducing Muse Your Personal AI Agent](#). September 8 2026 with later updates. Official launch announcement.
- [38] Mona Sarantakos with Christine Awad. [How We Designed Muse](#). September 2026. Primary design account.
- [39] Marcus Mendes for 9to5Mac. [Meta Muse Crosses Five Million Downloads amid Massive Advertising Push](#). September 30 2026. Reporting of Sensor Tower estimates.
- [40] Eric Hal Schwartz for TechRadar. [I Tried Meta Muse AI Agent](#). September 14 2026. Independent individual hands-on review.
- [41] Meta. [Security and Safety for AI Agents Our Approach with Muse](#). September 8 2026. Company engineering and safeguards account.
- [42] OpenAI. [Introducing Dots](#). September 29 2026. Official launch and availability statement.
- [43] Anthropic. [Cowork on Web and Mobile](#). July 7 2026. Official product announcement.
- [44] Anthropic Help Center. [Claude Cowork and Chat Are One Claude](#). Accessed October 5 2026. Current product and rollout guidance.
- [45] Tina Highfill and Jon D Samuels. [Early Evidence on the Relationship Between AI Costs and Prices](#). April 2026. BEA working paper with noncausal associations.
- [46] Soto Thieu and Allen. [The AI Buildout and the Economy](#). July 17 2026. Federal Reserve research note.
- [47] Chan and Shedania for ILO. [The Aggregation Paradox of AI](#). May 6 2026. Research brief and evidence synthesis.
- [48] International Energy Agency. [Key Questions on Energy and AI](#). 2026. Analysis and conditional projections.
- [49] Axios. [Anthropic Rolls Out First Brand Campaign](#). September 18 2025. Campaign reporting.

- [50] Anthropic. [Claude Is a Space to Think](#). February 4 2026. Company advertising and incentive commitment.
- [51] Anthropic. [Inviting Hard Questions](#). July 9 2026. Company campaign and engagement announcement.
- [52] Public Services Alliance. [AI Access Is More Than Availability](#). July 29 2026. Public-service access framework.
- [53] Public Services Alliance. [More on Pushback against AI Data Center Build Out](#). July 18 2026. Public-service commentary.
- [54] Public Services Alliance. [The Data Center Narratives](#). July 27 2026. Commentary and discussion of local public benefits.
- [55] Meta Help Center. [Muse Subscriptions](#). Accessed October 5 2026. Official free allowance and paid-plan guidance.
- [56] Meta Help Center. [Getting Started with Muse](#). Accessed October 5 2026. Official guidance on context boundaries and supervision.
- [57] Marc Andreessen. [The Techno Optimist Manifesto](#). October 16 2023. Investor advocacy and economic argument.
- [58] Public Services Alliance. [The Second Machine Age Revisited by ChatGPT](#). December 31 2024. Explicitly AI-generated book analysis and discussion.
- [59] Public Services Alliance reproducing Julian Horsey. [AI Meets Education Discover the Power of Google LearnLM](#). May 25 2025. Reposted product-oriented education article.
- [60] Public Services Alliance. [AI Interface Natives](#). August 5 2025. Community and interface commentary.
- [61] Public Services Alliance adapting Neaz Mujeri. [Building AI Literacy](#). November 7 2025. Adapted newsletter and training commentary.
- [62] Public Services Alliance reproducing David Shapiro. [The Elite Capture of AI](#). December 5 2025. Opinion about concentration and public ownership.
- [63] OpenAI. [Building Advertising for the Way People Use AI](#). October 5 2026. Company monetization announcement and independence commitments.

[64] OpenAI. [New Policy Ideas for the Intelligence Age](#). August 17 2026.
Policy research grant announcement.