

WW ANALYSIS LIMITED

UK-BASED ECONOMIC & POLICY RESEARCH

AI and the Value of University Education

Substitution, complementarity, and how universities should reposition

WW Analysis Limited · Research Report

Artificial intelligence can now perform much of what a university sells: explaining a concept, answering a question, drafting an essay, marking a problem set. This report asks which of those functions AI genuinely substitutes for, which it cannot substitute for and may even strengthen, and how universities should reposition around the difference.

01 · OVERVIEW

Executive Summary

Generative AI has been adopted by students faster than by almost any institution that teaches them. In the United Kingdom, 95% of undergraduates now report using AI in at least one way and 94% use it for assessed work, yet only 36% feel their institution encourages it and only 38% are provided with tools. That gap — near-universal informal adoption against partial institutional provision — is the central fact of the current moment, and it is where the risk sits.

The evidence on learning is now good enough to be directive rather than speculative, and it points consistently in one direction: unstructured access to a general-purpose model raises measured output while lowering learning, whereas the same underlying model, wrapped in pedagogy, produces large genuine gains. This is not a finding about AI. It is a finding about instructional design.

KEY MESSAGE

AI substitutes for the parts of university education that were already commoditised — content delivery, routine explanation, first-pass feedback. It does not substitute for credible assessment, supervised practice, or the networks and signals a degree confers. Universities that cut the second set to fund the first will erode the only part of their offering that AI cannot replicate.

- **Structure beats access.** In a randomised trial with roughly 1,000 high-school students, unguarded GPT-4 raised practice performance 48% but left students 17% worse than controls once it was withdrawn; a pedagogically constrained version of the same model raised practice performance 127% with no exam penalty.
- **The substitution is real but narrow.** What AI does well — scalable, patient, on-demand explanation — is precisely what large lecture courses did badly and expensively.
- **The complement is the credential and the network.** As verified output becomes cheap, verified *capability* becomes scarce, and the institutions that can certify it credibly gain rather than lose.
- **The labour market is repricing entry-level work, not degrees as such.** Graduate vacancies and early-career hiring have contracted sharply, which raises the return to demonstrable skill and lowers the return to credentials that signal only completion.

02 · INTRODUCTION

What is a university for, when the answers are free?

For most of the post-war expansion of higher education, universities held a near-monopoly on three things: expert explanation, structured practice with feedback, and a credential that employers trusted. Generative AI has broken the monopoly on the first, is competing seriously for the second, and has left the third intact but harder to defend.

The question this poses is not whether universities survive — the demand for what they certify is, if anything, rising — but which of their activities remain worth paying for. A university that understands itself primarily as a delivery mechanism for content faces a competitor with infinite patience and marginal cost near zero. A university that understands itself as an institution that develops and then credibly certifies capability faces a complement rather than a substitute.

This report takes the second view, but does not assume it. Sections 03 to 05 set out what the adoption and learning evidence actually shows; Section 06 identifies what survives substitution and why; Sections 07 to 10 examine the labour market, the economics of provision, and the regulatory constraints that shape what institutions can do; and Sections 11 to 13 set out the repositioning the evidence supports.

03 · ADOPTION

How fast, and by whom

Adoption has run ahead of institutional response at every level of education. Among UK undergraduates, the share using AI in at least one way rose from 66% in 2024 to 92% in 2025 and 95% in 2026. The share pasting AI-generated text directly into assessed work rose more slowly but steadily — 3%, then 8%, now 12% — which suggests that most use is assistive rather than substitutive, but that the substitutive tail is thickening.

Provider-side data tells the same story from the other direction. Analysis of ChatGPT usage finds that around 10% of all messages are tutoring or teaching requests, and that close to half of all messages from adult users come from users under 26. Education is not a niche application of these systems; it is one of the largest.

Institutional provision has begun to catch up, unevenly. The California State University system gave ChatGPT Edu to more than 460,000 students and 63,000 staff in early 2025, the largest single deployment anywhere; Oxford became the first UK university to extend it to all staff and students from 2025/26 after a pilot of roughly 750 people. Estonia has gone furthest at system level: its AI Leap programme puts a deliberately pedagogical assistant — one designed to guide reasoning rather than supply answers — into upper-secondary schools, expanding from September 2026 to vocational schools and a further 38,000 students.

Two features of this landscape matter for what follows. First, the tools students actually use are overwhelmingly general-purpose consumer models, not instruments built to teach. Second, institutional provision, where it exists, is mostly the same general-purpose model with a procurement contract attached. Very little of what is deployed at scale embodies the design features the evidence identifies as decisive.

04 · EVIDENCE

What AI does to learning

The research base has moved quickly from anecdote to randomised evidence, and the results are less contradictory than the headlines suggest. Taken together, the trials describe a single mechanism: generative AI reliably raises the quality of work produced with it, and the effect on learning depends entirely on whether the student's own cognitive effort is preserved or displaced.

Study and setting	Design	Finding
Bastani et al. (PNAS, 2025) Turkish high-school mathematics	RCT, ~1,000 students; unguarded GPT-4 vs pedagogically constrained tutor vs control	Unguarded: +48% with access, –17% on exam without it. Constrained tutor: +127% with access, exam performance level with control.
Kestin and Miller (Scientific Reports, 2025) Harvard introductory physics	RCT, 194 undergraduates; purpose-built tutor vs high-quality active-learning class	Students using the purpose-built tutor learned more than twice as much, in less time, with higher reported engagement.
World Bank pilot (2024–25) Edo State, Nigeria, secondary	RCT, nine public schools; paired use of GPT-4, teacher-supervised, six-week after-school programme	Gains of about 0.31 standard deviations in English, alongside better end-of-year exam performance.
Lei, Strömberg and Wu (working paper, 2026) China, ages 12–18	Panel of ~27,000 students over 30 months, nine subjects; homework, monthly closed-book exams, entrance exams	Homework scores +18% and time per assignment –30% (64 to 45 minutes); monthly exam scores –20% within six months, with entrance-exam performance also declining.

The pattern is consistent enough that the OECD's *Digital Education Outlook 2026* adopts it as a synthesis conclusion: where generative AI is neither built to teach nor used under structure, outsourcing tasks to it improves performance with no real learning gain. The mechanism has a name in the literature — metacognitive laziness — and a straightforward economic reading: the model substitutes for effort, and effort was the input that produced the learning.

Two qualifications deserve emphasis. The favourable trials all involve purpose-built systems, close supervision, or both, and none of them is the condition under which most students actually use AI. And the effect is distributionally uneven: survey evidence from seven European countries finds that 31% of students aged 12 to 17 use AI to obtain complete solutions, against 20% who use it to build personalised learning plans — with better-supported students more likely to fall in the second group. On present evidence the technology is at least as likely to widen attainment gaps as to close them.

INTERPRETATION

The relevant policy variable is not how much AI students use but whether the design of the task preserves the struggle that produces learning. That is a curriculum and assessment question, not a procurement question.

05 · SUBSTITUTION

What AI genuinely displaces

The functions most exposed to substitution are those that were already standardised, already delivered at scale, and already delivered indifferently. Recorded and lectured content delivery is the clearest case: a large first-year lecture transmits a fixed body of material to several hundred students at a fixed pace, and an AI system does the same thing at each student's own pace, on demand, without queueing. Routine tutoring — worked examples, the second and third explanation of a standard concept, unpicking a misunderstanding in a problem set — is similarly exposed, and here the substitution is welfare-improving: most students never had access to a patient one-to-one tutor at all.

First-pass formative feedback is the third area, and the most consequential for cost. Feedback on drafts, on code, on problem sets and on standard written work absorbs a large share of teaching time and is frequently rationed on grounds of workload. It is also the area where the productivity evidence is strongest on the staff side: in a randomised trial of 259 teachers across 68 English secondary schools, structured use of ChatGPT for lesson and resource preparation cut planning time by 31%, from 81.5 to 56.2 minutes per week. Sector data suggests the practice is already widespread — 44% of teachers in England reported using generative AI for school activities in 2024/25, concentrated in preparation rather than live teaching or marking. Analysis of educators' use of Claude points the same way: curriculum development accounted for 57% of the conversations examined, against 7% for assessing student performance.

Finally, and least comfortably, unsupervised written assessment has been substituted — not in the sense that it can be done better by AI, but in the sense that it no longer measures what it was designed to measure. This is the substitution that has forced the fastest institutional response, and it is treated in Section 08.

What unites these four is that none of them was a source of durable advantage. They were the cost centres of teaching, not its value. Their displacement releases resources; it does not in itself threaten the institution.

06 · COMPLEMENTARITY

What survives, and why it becomes more valuable

The parts of university education that AI cannot substitute for share a common structure: their value depends on something being scarce, verified, or embodied in a relationship. Cheap generation does not erode these; it raises their relative price.

Credible certification

A degree is valuable to an employer to the extent that it carries information the employer cannot cheaply obtain otherwise. When written output becomes free, the informational content of unsupervised written work collapses — and the value of assessment that cannot be outsourced rises correspondingly. Invigilated examination, oral defence, supervised practical work and observed performance are, in this frame, not nostalgic retreats but the assets whose scarcity has just increased. Universities are among the few institutions with the standing to certify capability in a way labour markets accept, and that standing is a genuine moat.

Peer and professional networks

The network a student joins is rival, cumulative and impossible to generate synthetically. Cohort effects, supervision relationships, alumni access and the professional judgement acquired by working alongside practitioners are all forms of capital that exist only through co-presence over time. Notably, the same student survey that records near-universal AI adoption also records around 15% of students using AI for companionship or to address loneliness, with the effect on loneliness split almost evenly between those who feel less lonely (21%) and more lonely (20%) — a reminder that AI is not a substitute for the social good universities supply, and may in some cases substitute for it badly.

Hands-on research and supervised practice

Laboratory work, fieldwork, clinical placement, studio practice and genuine research apprenticeship all require access to equipment, patients, sites, data or expert judgement that no model provides. These are also the settings in which tacit knowledge — how to recognise that a result is wrong, how to frame a question worth asking — is transmitted, and they are correspondingly expensive. The strategic implication is uncomfortable: the activities that are most defensible are also those with the worst unit economics.

The productive struggle itself

The trial evidence in Section 04 implies that difficulty, correctly calibrated, is not a defect of instruction but its active ingredient. An institution's capacity to design tasks that keep effort in the right place — and to hold students to them — is a form of expertise that becomes more valuable precisely as the temptation to outsource effort becomes universal.

07 · LABOUR MARKET

The graduate premium under pressure

Any assessment of the value of university education has to reckon with what is happening at the point of entry to the labour market, where the deterioration has been sharp and recent. In the United Kingdom, advertised graduate vacancies fell to 8,383 in July 2026, the lowest in a decade of tracking and 45.6% below the 15,397 advertised a year earlier, with competition rising to 2.14 jobseekers per vacancy from 1.93. Separate data recorded roles advertised for recent graduates down about a third year-on-year in mid-2026. In the United States, unemployment among recent graduates stood at about 5.6% through the second quarter of 2026 against roughly 4.2% for all adults — an unusual inversion of the normal graduate advantage — with underemployment at about 42%.

Whether AI is the cause is contested, and the honest answer is that attribution is not yet settled: employer national insurance costs, minimum-wage increases and a weak cycle all coincide with AI adoption in the same period. The most careful available evidence is suggestive rather than conclusive. Tracking of US payroll microdata finds employment of 22- to 25-year-olds in the most AI-exposed occupations roughly 19% below where it would have been had it kept pace with similarly aged workers in less-exposed occupations, up from about 15% a year earlier, with no comparable effect for older workers in the same occupations. A concentration of losses in early-career, high-exposure work is what AI substitution for junior tasks would look like; it is also what a hiring freeze that falls first on new entrants would look like.

Student behaviour has already begun to respond, and not always in the direction the narrative predicts. Enrolment in computer and information science in the United States fell 8.1% in 2025/26, the steepest decline of any field, at the same time as dedicated AI degree programmes proliferated — a migration within technical education rather than away from it. Whether this is a rational reallocation or a lagged overreaction to a single bad hiring cycle will not be clear for several years.

STRATEGIC READ

The repricing is concentrated in entry-level tasks, not in graduate capability. That shifts the return away from credentials signalling completion and towards those evidencing demonstrable, verified skill — which is an argument for changing what a degree certifies, not for doubting whether certification matters.

08 · REPOSITIONING

Assessment first, because it is already moving

Assessment is where institutions have responded fastest, because the alternative was to keep awarding marks for work of unknown provenance. Nearly two-thirds of UK undergraduates — 65% — report that assessment has changed significantly in response to AI. In the United States, 57% of faculty at public universities report changing assessment strategy because of AI, and of those, 38% have reintroduced handwritten examinations. Princeton will require all in-person undergraduate examinations to be supervised from 1 July 2026, ending a 133-year tradition of unproctored testing; the University of Bath moves to time-limited, invigilated, in-person closed examinations from 2026/27.

The reflex to invigilate is defensible but insufficient on its own, for two reasons. It restores the integrity of the measurement while narrowing what is measured — handwritten closed-book examination is a poor instrument for the extended, iterative, evidence-marshalling work that most graduate employment actually involves. And it addresses provenance without addressing capability: a student who has outsourced two years of formative work will fail a supervised examination, which protects the credential but does not educate anyone.

The more durable response combines three moves. Assess the process alongside the output, so that drafts, decisions and revisions carry marks. Make AI use explicit and assessable, requiring students to critique model output rather than pretending it is absent. And concentrate summative weight on the settings AI cannot enter — supervised, oral, practical and performed. Institutions attempting all three simultaneously are, unsurprisingly, finding it expensive; the workload evidence in Section 05 is where the offsetting resource has to come from.

One risk deserves naming. Students report anxiety about false accusations of academic misconduct, and AI-detection tools remain unreliable enough that this anxiety is rational. An enforcement regime built on probabilistic detection rather than assessment redesign will generate injustice at scale, and will do so disproportionately among students writing in a second language.

09 · CAPABILITY

What the AI era actually demands

Students are clear about what they want and equally clear that they are not getting it: 68% believe AI skills are essential to thrive, while fewer than half — 48% — feel teaching staff are helping them develop those skills, with arts and humanities students least well served. The gap is not primarily about tool training, which students acquire unaided. It is about the judgement that determines whether the tool helps or harms.

Three capabilities follow directly from the evidence in this report. The first is verification: the ability to establish whether a fluent, confident, plausible output is actually correct — which requires domain knowledge that cannot itself be outsourced, and which is why AI literacy cannot be taught as a standalone module. The second is problem formulation: as execution becomes cheap, the binding constraint moves upstream to deciding what is worth executing and specifying it precisely enough to be useful. The third is knowing when not to use it — the metacognitive discipline to recognise that a task is one where the struggle is the point.

This has a structural implication for curriculum design. If verification requires domain expertise, then the AI era raises rather than lowers the return to deep disciplinary knowledge, and the widespread instinct to replace subject depth with generic digital skills gets the lesson backwards. China's programme restructuring — 12,200 undergraduate programmes revoked or suspended between 2021 and 2025, with capacity redirected towards AI and adjacent fields, against a record 12.7 million graduates in 2026 — represents the boldest test anywhere of the opposite bet, and its results will be informative whichever way they fall.

10 · CONSTRAINTS

Finances and regulation

Repositioning has to happen inside two binding constraints. The first is financial, and in England it is severe: more than a third of higher education providers — 35.8% — reported a deficit in 2024/25, with the regulator forecasting 42.7% for 2025/26, against non-UK entrants down 7.7% and 9.0% below providers' own forecasts, and an international student levy projected to cost the sector around £570 million in 2028. The temptation this creates is precisely the wrong one: AI offers visible savings in exactly the activities that are cheap to cut and hard to defend cutting — contact hours, feedback, small-group teaching — while the defensible activities carry the highest unit costs. An institution that uses AI to reduce contact while raising fees is optimising itself into the substitutable half of its own business.

The second constraint is regulatory. Under the EU AI Act, AI systems used for admissions and access decisions, for evaluating learning outcomes, for determining appropriate levels of education, and for monitoring prohibited behaviour during tests are classified as high-risk under Annex III, with obligations for standalone high-risk systems in these Annex III sectors deferred to 2 December 2027. The Act's separate AI-literacy duty is already operative, and applies to staff. In England, the Department for Education has moved from guidance to safety standards that generative AI products should meet for use in educational settings, with the expectation that staff check output for accuracy, bias, age suitability and curriculum fit before it reaches learners.

The practical reading is that the compliance burden falls hardest on exactly the high-stakes applications — admissions, grading, proctoring — that institutions under financial pressure are most tempted to automate. Deploying an assistant to help students learn attracts light obligations; deploying one to decide who gets in, or what mark they receive, attracts heavy ones, and on the current evidence about bias and reliability, that asymmetry is well judged.

11 · SCENARIOS

Three paths

The following stylised paths are illustrative rather than predictive, but they capture how the choices in Sections 08 to 10 interact.

Path	What the institution does	Likely outcome
Cost substitution	Uses AI to reduce contact hours, feedback and small-group teaching while holding fees; leaves assessment broadly unchanged	Short-term margin relief; the credential's informational content erodes and the institution competes directly with a free substitute
Prohibition	Bans AI use, polices it through detection tools, restores closed examinations without redesigning what is taught	Integrity partially restored; graduates enter an AI-saturated labour market without the judgement to use it, and false-accusation harm accumulates
Redesign	Reinvests substitution savings in supervised, oral and practical assessment; makes AI use explicit and assessable; protects disciplinary depth as the basis of verification	Higher unit costs and slower change, but the credential and the capability it certifies both hold their value

The distance between the first and third paths is not primarily technological. Both are available to any institution; they differ in whether savings from substitution are taken as margin or reinvested in the complement. That is a governance choice, and financial distress pushes hard towards the first.

12 · IMPLICATIONS

For universities and policymakers

- Treat the design of AI use as a curriculum decision, not a procurement decision.** The trial evidence shows the same model producing opposite learning outcomes depending on how the task is structured; buying licences settles almost nothing that matters.
- Reinvest substitution savings in the complement.** Efficiency gains in preparation and first-pass feedback should fund supervised, oral and practical assessment — the activities whose scarcity value has risen — rather than being taken as margin.
- Redesign assessment rather than relying on detection.** Probabilistic detection tools generate injustice at scale and do not restore the informational content of the credential; process-based, supervised and performed assessment does.
- Protect disciplinary depth.** Verification of AI output requires domain expertise, so the AI era raises the return to deep subject knowledge; replacing it with generic digital skills inverts the lesson.
- Close the support gap deliberately.** With 68% of students calling AI skills essential and 48% feeling supported, the binding constraint is staff capability and time — which requires funded training, not exhortation.
- Regulate high-stakes uses, not learning uses.** Admissions, grading and proctoring warrant the heavier obligations the EU framework imposes; assistive learning tools do not, and treating them alike will suppress the beneficial case while doing nothing about the harmful one.
- Monitor distributional effects explicitly.** Present evidence suggests better-supported students use AI as a tutor and less-supported students use it as a shortcut, which makes attainment-gap monitoring a first-order requirement rather than an equity afterthought.

13 · CONCLUSION

The commoditised half and the scarce half

Universities do two things that have always been bundled together and priced as one: they deliver instruction, and they develop and certify capability. Generative AI has unbundled them. The delivery half is now available at near-zero marginal cost, at higher patience and better availability than most institutions ever managed. The development and certification half has become scarcer and more valuable in exactly the same movement, because the cheapness of output is what makes verified capability worth paying for.

The strategic error available to universities is to defend the half that has been commoditised — by restricting AI, by competing on content delivery, or by using AI savings to thin out the contact and assessment that constitute the other half. The opportunity is the reverse: to accept substitution where it has occurred, to take the productivity gain, and to spend it on supervised practice, credible assessment, and the disciplinary depth that makes a graduate able to tell whether the machine is right. On the evidence assembled here, that is not a defensive adaptation. It is a stronger position than the one the sector occupied before.

APPENDIX A

Human capital, signalling, and why the distinction now bites

Economics offers two accounts of why education raises earnings, and the AI transition separates them in a way that ordinary times do not. On the human-capital account, education raises earnings because it makes people more productive. On the signalling account, it raises earnings because completing it reveals pre-existing ability to employers who cannot observe that ability directly. Most empirical work finds both mechanisms operating, and for practical purposes the distinction has rarely mattered much: whichever was doing the work, the prescription was similar.

Generative AI drives a wedge between them. If a student outsources coursework to a model, the human capital is not accumulated — the trial evidence in Section 04 shows exactly this — but the signal is still emitted, because the credential is awarded. Cheap, high-quality output therefore contaminates the signal without adding capability, and it does so silently. Employers respond to a degraded signal by discounting it and by seeking substitutes: work-sample tests, technical interviews, trial periods, portfolios. Part of the graduate-hiring deterioration in Section 07 is plausibly this discounting rather than pure labour substitution, though the data cannot currently separate the two.

The corollary is that assessment integrity is not a matter of institutional propriety but of market infrastructure. A university that can credibly certify what a graduate can do without assistance supplies information that has just become scarce and valuable. One that cannot is left selling a signal that employers have learned to disregard — and, on the human-capital side, an education that a free tool has partially replaced.

APPENDIX B

Notes on the evidence

- **Design.** The four studies in Section 04 are the strongest available identification: three are randomised controlled trials, and the fourth is a 30-month panel with within-student variation across nine subjects and both low- and high-stakes assessment.
- **Generalisability.** Settings range from Turkish and Chinese secondary schools to Harvard undergraduates and Nigerian public schools. The convergence of the mechanism across such different contexts is the main reason for treating it as directive; the magnitudes are not transportable.
- **What is not yet known.** No study yet tracks a full cohort from AI-saturated secondary education through higher education into the labour market, so the cumulative effect on human capital is unmeasured. Attribution of graduate-hiring weakness between AI substitution, signal discounting and cyclical factors also remains open.
- **Survey caution.** Student-reported perceptions of AI's effect on their own learning are poorly correlated with measured learning — the trial evidence shows students rating unguarded AI highly while learning less from it — so self-report is used here for behaviour and provision, not for outcomes.

GLOSSARY

Key terms

Term	Meaning
Substitution	AI performing a function previously supplied by the institution, at lower cost or higher availability
Complementarity	AI raising the value of a function the institution supplies, rather than replacing it
Metacognitive laziness	The displacement of a learner's own cognitive effort by an external tool, raising output while reducing learning

Term	Meaning
Formative assessment	Assessment intended to guide learning in progress, as distinct from summative assessment, which certifies attainment
Signalling	The mechanism by which a credential conveys information about pre-existing ability to employers who cannot observe it directly
High-risk AI system	Under the EU AI Act, a system in a listed Annex III use case — including education admissions, outcome evaluation and test proctoring — subject to enhanced obligations

FURTHER READING

Sources and context

Learning effects: H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakci and R. Mariman, “Generative AI without guardrails can harm learning: evidence from high school mathematics”, *PNAS*, 2025; G. Kestin and K. Miller, “AI tutoring outperforms in-class active learning”, *Scientific Reports*, June 2025; V. Lei, D. Strömberg and Y. Wu, working paper on generative AI and student learning in China, 2026; World Bank, generative-AI tutoring pilot, Edo State, Nigeria, 2024–25; OECD, *Digital Education Outlook 2026*, January 2026.

Adoption and provision: HEPI and Kortext, *Student Generative AI Survey 2026* (HEPI Report 199, March 2026, 1,054 UK undergraduates surveyed by Savanta in December 2025); A. Chatterji et al., “How People Use ChatGPT”, NBER Working Paper 34255, 2025; Anthropic education reports on student and educator use of Claude; Ipsos for the Vodafone Foundation, survey of 7,000 students aged 12–17 across seven European countries; Education Endowment Foundation and NFER, *ChatGPT in Lesson Preparation* teacher-choices trial; Estonian Ministry of Education and Research, AI Leap programme.

Labour market: Stanford Digital Economy Lab, “Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence”, August 2026 update; Adzuna graduate vacancy series, July 2026; Federal Reserve Bank of New York, *The Labor Market for Recent College Graduates*, 2026.

Institutions and regulation: Office for Students, *Financial sustainability of higher education providers in England: 2026*; Regulation (EU) 2024/1689 (EU AI Act), Article 4 and Annex III; UK Department for Education, generative AI standards for education; International Center for Academic Integrity, faculty assessment survey; institutional announcements from Princeton University and the University of Bath.

About WW Analysis Limited. WW Analysis is a UK-based economic and policy research consultancy specialising in the economics of technology, labour markets and public policy. For bespoke analysis of AI adoption, education and skills policy, or the labour-market effects of automation, contact ww.analysis.uk@gmail.com.