

TAMING THE CROCODILE

FUTURE-PROOFING SCHOLARLY PUBLISHING IN THE AI-FIRST ECOSYSTEM

Sneak peek at phase 1 report and phase 2 plans,
for those considering joining the project



AUTHOR

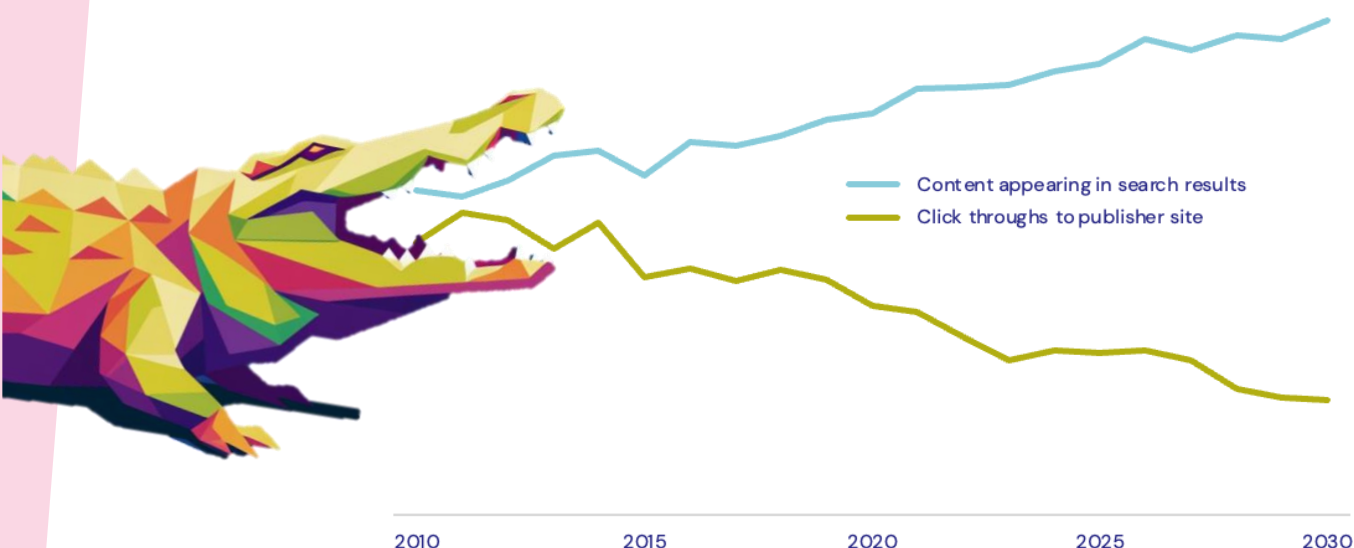
Charlie Rapple

Co-founder and Research Lead • Kudos • charlie@growkudos.com

 **Kudos**
Science needs Stories

WHAT IS THE CROCODILE AND WHY DO WE NEED TO TAME IT?

The Crocodile Effect¹ is a rapidly growing threat to the reach and readership of scholarly and scientific content. It is the term used to describe the gap between impressions of your content in search engines, and click throughs to the full text on your platform.



This issue, also called “zero-click search”, has grown hugely since the growth of AI Overview panels in general search engines, and the whirlwind adoption of gen AI – between them, these two forms of AI-mediated discovery have become the starting point for over a third of information seeking journeys, and reduced usage of content by up to 70%.

AI discovery threatens not only usage, but also research integrity and transparency. It takes content out of context (divorcing it from quality indicators and trust signals), and amplifies / camouflages misinformation (treating non-authoritative user-generated content the same as peer-reviewed content).

The “Taming the Crocodile” initiative launched in late 2025 to collaborate in exploring and addressing this challenge. Phase I has been sponsored by 18 organizations including major publishers, societies and platform / service providers. It set out to understand how AI is changing the discovery of, and engagement with, scholarly content – and what that means for publishers and platforms. It comprised 2 surveys (users’ and librarians’ awareness / understanding / expectations of AI in information discovery) and an environment scan (over 300 sources + 10 events) to identify quick wins and medium / longer-term actions.

This document is for people interested in buying a copy of the Phase I report and / or participating in (sponsoring) Phase II. It includes a small sample of the findings and recommendations, sets out what is covered in the full Phase I report, and explains what Phase II will comprise. I’ve included pricing and contact details at the end for those wishing to proceed. Phase II will be completed during Q3–4 2026.

¹ The phrase ‘crocodile effect’ derives from a LinkedIn post by SaaS marketing expert Tim Soulo: https://www.linkedin.com/posts/timsoulo_the-past-12-months-of-gsc-data-for-the-ahrefs-activity-7341034461481902080-JgxY/



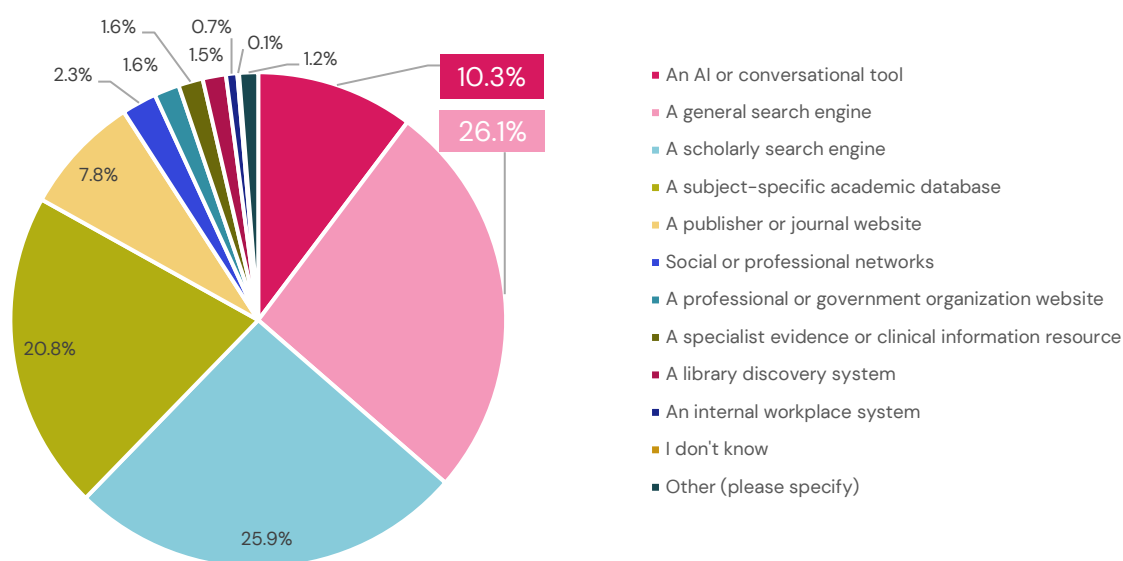
SNEAK PEEK AT SOME PHASE I FINDINGS & RECOMMENDATIONS

AI has captured a third of research discovery.

One in ten researchers now turn directly to AI tools as the starting point for their information seeking; a further 25% start in general search engines where AI-generated answers are now embedded. This extraordinarily rapid shift is even more pronounced among early-career and corporate researchers (twice as many start with an AI tool).

When you need to find information online, where do you most often start your search?

(n = 11,416)



What should publishers and platforms do about this?

"Search Engine optimization" is now "Search Everywhere Optimization" and requires new tactics. The single most frequent recommendation from AI Optimization agencies is to develop an AI discovery layer of structured, "answer-ready" plain-language summary content ("bottom line up front" – what is this about and why is it important). What was a "nice to have" is now essential if you want to reclaim / rebuild traffic, and retain brand visibility and citations. Content must be present in server-rendered, semantic HTML (AI crawlers don't execute JavaScript). Accurate listings in Semantic Scholar, OpenAlex and (where relevant) PubMed Central are bare-minimum requirements for the indexes AI systems draw on. The same principles apply in China, redirected at Baidu, DeepSeek and Doubao.

The full report includes guidance on what kinds of content will help increase visibility in AI, how to structure it and mark it up, how to make sure your SEO / GEO / AEO settings are not hindering discoverability, and further details on what you can do to optimize discovery by users in China.

Longer-term suggestions include semantic enrichment and even changing the structure / layout / flow of articles to more prominently display the signals that AI (and indeed, humans) are looking for.

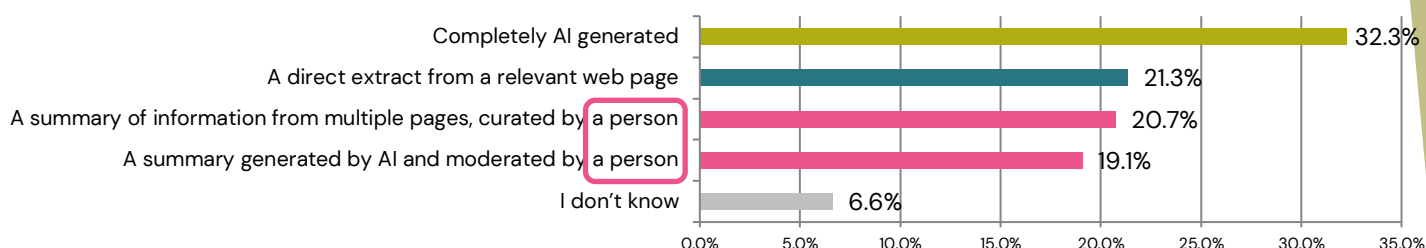


Users don't always know they are using AI.

Alarmingly, 40% wrongly think AI overviews generated on the fly in search engines have been created or curated by a human. That figure (46%) was higher among students. Librarians report users routinely requesting articles that do not exist, because of hallucinated responses by AI.

To the best of your knowledge, what is the source of information in knowledge panels?

(n = 9,674)



What should publishers and platforms do about this?

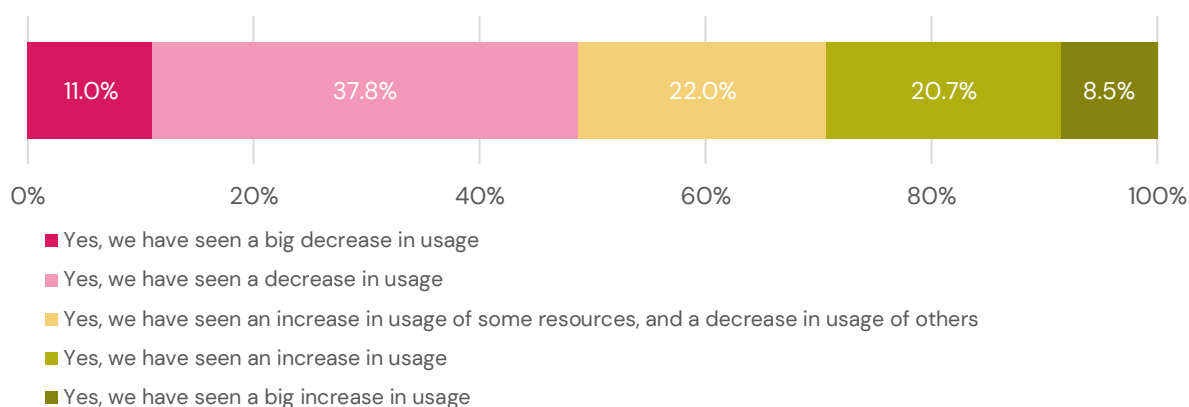
The report gives suggestions for how to support librarians in improving users' AI literacy through co-development of AI educational content, with suggested topics and formats.

Usage is dropping, and putting subscriptions at risk.

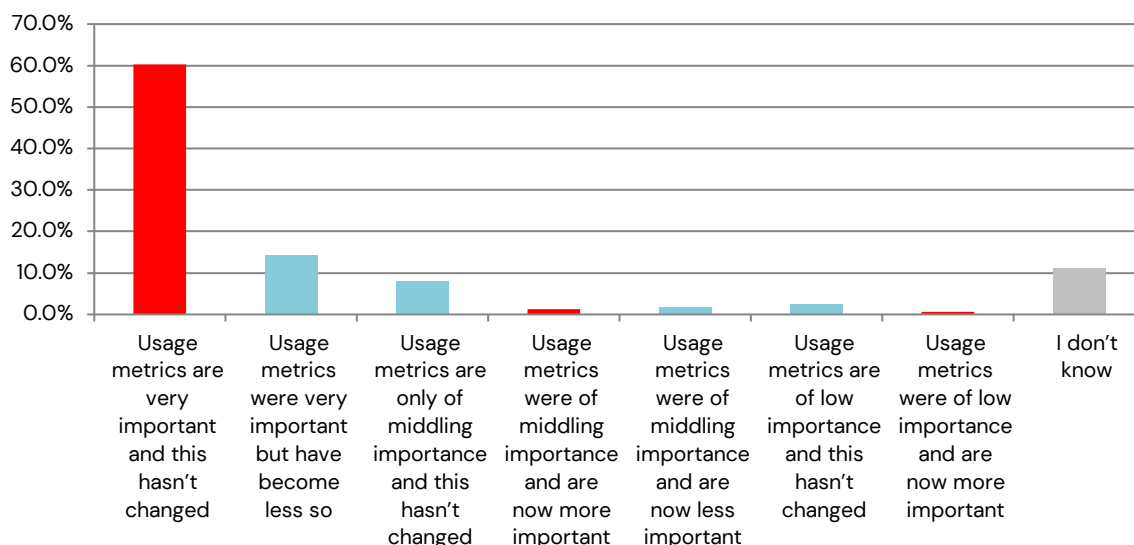
Librarians are reporting substantial changes in researchers' behaviour. Of those who have noticed a change in usage of primary sources since AI became more prevalent in information discovery, **almost half report a drop in usage of primary sources due to AI discovery.**

Although there are other ways to conceptualise value, usage metrics are still driving purchasing decisions. In our priority actions we provide near-term practical steps and long-term guidance on how publishers and platforms can retain their current value / position in this ecosystem

Have you noticed changes in usage of the library's websites and resources since AI has become more prevalent in information discovery? (n = 82)



How important are traditional usage metrics (downloads, clicks, sessions) in your purchasing decisions today? (n = 161)



What should publishers and platforms do about this?

Publishers need to maximize discoverability of content in AI resources (as above) to help 'fill the top of the usage funnel' – but must also work with librarians to collectively agree new definitions of 'what counts', so that the full gamut of 'usage' is being reported and understood. COUNTER is working on new definitions to enable publishers to report off-platform usage including in AI tools. The challenge is ensuring that that usage is being properly recorded by the AI companies. In Taming the Crocodile's Phase II, we will be working to agree some best practices around discoverability metadata to ensure we are consistent in how we share information with AIs and what we ask from them in return. Building the relationships with the AI companies at this operational / technical level helps provide a nexus for the broader discussions around how 'usage' should be captured and reported. This also dovetails with NISO's work around attribution and provenance. On which note:

Publishers need to re-position as the gatekeepers of trust.

For AI companies to develop sustainable business models, they need to be able to guarantee accuracy and authority of the responses they generate. They will increasingly look to scholarly publishers as valued partners that have well-established and proven workflows for creating and validating trust. Plus, librarians want publishers to play an active role in ensuring AI systems accurately represent scholarly content; they want transparency about publishers' engagement with AI, and have insightful views on how content should be licensed for AI training or agentic querying. Substantial sections of the Phase I report are devoted to findings and recommendations around how we can better mark-up and communicate signals of trust.



WHAT DOES THE FULL PHASE 1 REPORT CONTAIN?

Findings, analysis and recommendations from two surveys (11,500 researchers and 200 librarians) and an environment scan of more than 300 sources.

- **Easy to read:** full data and findings provided but also digested into **4pp executive summary of the most unexpected / important takeaways, written for non-experts**
- **Easy to action:** Detailed recommendations digested into **60 priority actions** for scholarly and scientific publishers, platforms and content providers – frankly, anyone with a website.
- **Easy to delegate:** Actions are grouped for marketing, editorial, customer service, account management and technology teams, and also provide points of consideration for your longer-term strategic response.

QUOTES FROM PHASE I SPONSORS:

"It's a genuinely impressive piece of work, and the whole engagement has been a pleasure from start to finish. The finding that matters most is that researchers rely on AI but do not trust it, and that evidence and clear attribution are what earn both the click and the citation. The insights are exactly what we hoped for, and we are already using them to shape how we help authors make their work discoverable and credible in an AI-mediated world."

*Shane Rydquist, Associate Vice President,
Delivery and Solutions at Cactus Communications*

"The findings give us all such a strong foundation to help adapt our industry infrastructure."

Emilie Delquié, Chief Product & Customer Success Officer at Silverchair

"The fast-moving and unpredictable impact of AI on all aspects of scholarly communication makes it essential to stay ahead. We all need to act quickly, and responsibly, on the findings here to ensure that we continue to support researchers, the wider community, and the integrity of scholarly communication."

Dan Penny, Head of Market Intelligence at Springer Nature

1. INTRODUCTION: GET-UP-TO-SPEED SECTION COVERING:

- When did AI start to be 'a thing' in scholarly information discovery?
- How and where do users encounter AI in information discovery now?
- How and why does AI ingest scholarly and scientific content?
- How does AI generate overviews?
- Why has all this become a problem?
- Are there potential positives?



2. ENVIRONMENT: WHAT ARE PEOPLE DOING ABOUT ALL THIS?

Insights drawn from a range of sources within and beyond the scholarly communications sector, including digital marketing consultancies, consumer retail businesses, corporate / healthcare / pharmaceutical information businesses, reference platforms, news publishers, legal and financial information providers and others.

- All the Os: SEO, GEO, AEO, AIO and back to SEO
- Structured summaries: plain language, question-oriented
- Auditing and optimizing your search / indexing settings
- Licensing use of your content
- Content enrichment, structure and tagging
- What to do in China

3. UNIQUE INTELLIGENCE: WHAT DO CUSTOMERS EXPECT?

Researcher survey: “How You Search: AI and Knowledge Tools in Your Professional Work”

- **Sizeable sample of quality responses:**
11,416 responses after quality filtering
- **Broad geographic representation:**
Asia 35.3%; Europe 28.6%, Africa 12.2%, N. America 11.1%, S. America and Latin America 4.7%, Australasia & Pacific 4.5%, Middle East 3.7%.
- **Valid sample sizes in all subject areas:**
26.2% Medical and Health Sciences, 13.5% Life Sciences, 13.3% Engineering & Technology, 4.4% Business & Management (full breakdown below). Over 100 respondents in every field except the Arts (n = 30).
- **Wide range of ages and roles:**
73.1% academic researchers; 10.3% students; 6.2% healthcare practitioners; 2.3% corporate researchers
- **Information discovery:**
 - Where do users start their information discovery? Which discovery channels do users go to for which types of content? How is this changing?
- **AI Overviews**
 - Do users know AI when they see it? Do they trust it? In which circumstances do users still click through, and what is their intent when they do so?
- **Expectations of publishers**
 - Do users notice publisher / journal brands in discovery? Does this increase trust? Do they think publishers should have more or less control over how content is displayed in AI?



Librarian survey: “AI in Libraries: Adoption, Impact, and Challenges”

- **Good sample size:**
195 quality-filtered responses
- **Global insights:**
N. America 43.5%, Europe 27.3%, Asia 14.3%, Africa 4.5%, Australasia 3.2%, S America 1.3%, Middle East 0%
- **Range of institution types:**
28.9% research-intensive, 32.1% teaching-focused, 25.2% dual-focus; 3.1% healthcare; 2.5% government; 1.9% corporate
- **Learnings from all levels of seniority:**
25.8% library directors; 16.4% collections; 15.1% subject librarians; 11.3% systems librarians; 10.1% research support
- **Librarians’ perceptions, engagement and choices**
 - Is AI a threat or an opportunity? Are libraries engaging, and how / where? Which tools are librarians using / paying for? How do they feel about publisher / vendor AI features? Where is the skills / education gap that publishers could help to fill?
- **Users’ behaviour, engagement and choices**
 - What tools are being used by / licensed for libraries’ patrons? What changes are librarians seeing, and what are they worried about? Are they seeing a decline in usage of publisher content, and is this going to affect purchasing decisions? What metrics do librarians want publishers to provide?
- **What do libraries expect publishers and platforms to do in terms of AI access to content for training and querying?**

4. URGENT ACTIONS: WHAT DO PUBLISHERS NEED TO DO?

- Create AI-optimized discovery content layers
- Test, benchmark and monitor AI visibility
- Develop machine-readable signals of EEAT (Experience, Expertise, Authority, Trust)
- Partner with librarians to educate users around AI literacy
- Ensure adherence to basic SEO best practice, and add AIO / GEO techniques
- Implement entity SEO and semantic enrichment
- License phase 1 report / participate in phase 2 to receive many more detailed recommendations.
- And more – 60 recommendations in total



WHAT WILL PHASE 2 TACKLE?

1. PUBLISHER BENCHMARKING: Q3 (EARLY REPORT IN OCT '26)

Survey to anonymously (in 'escrow') gauge usage changes (growth as well as loss), and awareness / progress around recommendations from the phase 1 environment scan. No pressure on sponsors to provide any data but an opportunity to benchmark your organization against wider progress in the sector. We will capture high-level intelligence from publishers on points such as:

- The level of impact zero-click search has had on usage of their content
- Whether they are blocking or enabling AI crawlers and agentic bots
- Whether they have undertaken any specific AEO / GEO or AI-metadata creation
- What they are doing to ensure their content is represented in open knowledge graphs
- Whether they have a strategy for content licensing
- The extent to which they are consulting or co-developing with libraries or corporate customers

Publishers, platforms and providers won't be able to gather this information independently but using Kudos as a neutral third-party enables you to understand where your organization sits in terms of AI-readiness, and what you need to do to maintain or strengthen your competitive edge in this area.

2. USER OBSERVATION: Q4 (EARLY REPORT IN NOV '26)

- Observation of users in different fields, roles, countries and at different career levels
- Watching what they do, in order to identify / probe for answers to questions such as:
 - What triggers information seeking, and how do they choose where to start?
 - To what extent have they moved from keyword to conversational search?
 - How do they decide when they have information they need from a source they trust?
 - How aware are they of AI during their journey and how consciously are they making decisions about trust?
- Analysis, cross-tabulation and guidance as to how to put the findings into practice.

This data and analysis will allow publishers, platforms and providers to better understand exactly how researchers are integrating these tools, building on the Phase I findings, but also revealing subconscious preferences, behaviors, and priorities. This mapping out of user journeys will enable us to make further recommendations for creating AI-optimised content, re-capturing audience attention, and improving click-through rates to your primary sources



3. COLLABORATIVE DEVELOPMENT OF A CODE OF PRACTICE / SCHEMA TO STANDARDIZE AEO / GEO METADATA (Q4: EARLY DRAFT IN DEC '26)

At May's [COUNTER / NISO / Cambridge University Press workshop](#) the focus was on provenance, attribution and usage metrics in AI-mediated discovery, and within that context there was useful discussion about what the scholarly & scientific publishing sector *wants* from AI companies, in terms of what data we want them to ingest and use, and what metrics we want them to capture and share. There is recognition that working together as a sector and defining a consistent 'ask' gives greater potential for that ask to be met (and indeed, the AI companies represented in the room, and consulted in other contexts for this study, largely had the attitude "tell us clearly what you want and we will see what we can do").

The same applies around discoverability: the norms governing how AI companies ingest, represent and attribute content are being shaped and set right now. We need to move quickly to ensure our expectations are heard and met.

What kinds of content / metadata do we want AI companies to ingest and use (to better surface and represent scholarly and scientific content in AI-mediated discovery), and how should we structure that content and mark it up.

Following consultation with NISO, COUNTER and STM, we have agreed to use phase 2 of the crocodile project to develop a first iteration of best practices / standardization of discoverability metadata. The goal in the near term is to agree content types and metadata properties. This shared framework will give us a clear, consistent basis for dialogue with AI companies about what we provide and what we expect in return. The longer-term goal is for this work to form the basis of standardization through the appropriate bodies. I see the role of the Crocodile project as similar to when I set up UKSG's KBART working group, which went on to become a NISO Recommended Practice. With limited capacity in the standards organizations and several fronts on which progress needs to be made, we can use Phase II of the Crocodile to do the groundwork on the discoverability piece which is an essential precursor / complement to work around provenance, attribution and usage metrics.

SPONSORS ALREADY COMMITTED TO PHASE II

Include Springer Nature, Wiley, Elsevier, Silverchair, IEEE, and American Physical Society.



HOW TO PARTICIPATE IN PHASE II

- **Buy the Phase I report:** £15,000 + VAT (17.750€ / \$20,000)
 - Custom presentation by Charlie Rapple to your organization, to bring your team up to speed
 - Slide decks and graphics for your use in briefing colleagues and customers
 - 113 page report
 - 10 page summary of findings and recommendations
 - Full export of data from both surveys
 - Summary charts from both surveys
 - Full 22 page list of 311 references all linked for easy review
 - 7 page glossary explaining all concepts and terms in plain language
 - Non-exclusive, royalty-free licence to quote from and use content for internal and external purposes i.e. in customer presentations, company updates (with a fair use approach – not reproducing the report in a way that undermines its value to Kudos and others). All intellectual property in the research methodology and outputs remains with Kudos.
- **Buy Phase I report and join Phase II:** £25,000 + VAT (29.500€ / \$33,500)
 - All Phase I report features plus:
 - Immediate involvement in Phase II, kicking off in August 2026
 - Participation in project kick-off and monthly update calls
 - Regular update emails sharing findings and recommendations throughout the project (for competitive advantage through early insight and actions rather than waiting for the final report)
 - Brand recognition in promotional materials before, during and after the project
 - Full report, analysis and recommendations (findings from publisher survey, user observation, recommendations around AIO / GEO content and metadata)
 - Survey results exports (basic dataset and summary graphs)
- **Buy phase 1 report, join Phase II as a headline sponsor:** £30,000 + VAT (35.250€ / \$40,250)
 - All of the above plus opportunity to steer research scope / instruments / schema etc.
 - Custom presentation to your team of Phase II findings and recommendations by Charlie Rapple.

Contact charlie@growkudos.com to sign up or arrange a discussion.



WHO SPONSORED PHASE I?

