

Early career researchers, scholarly communications and generative AI Schedule to Code H3-P1 2024-01-12

Section X (Background/demographics) was labelled section A in H2-R3, I have relabelled the questions so that the other sections retain the section numbers used for the interview schedule. Questions and formats are unchanged from H2-R3.

H3-Q Permitted codes	H3-P1 Question Background/ demographics	R3-Q
X01 CC0nnSUBX	ECR ID	A01
X02 H1, H2	H1, H2?	A02
X03 M,F	Gender (M,F)	A03
X04 YY-MM-01	DoB [Year- Month-01]	A04
X05 ISO 3166 country code	Working Country Date of Interview	A05
X06 YY-MM-DD	[Year-Month-Day]	A06
X07 CN, ES, FR, MY, PL, RU, GB, US	Cohort Country Nationality/ citizenship (if different)	A07
X08 ISO 3166 country code	Dependants/ carer?	A08
X09 Y, N, NA	Other countries in which worked as a researcher.	A09
X10 ISO 3166 country code	Broad disciplinary category:	A10
X11 MATH, PHY, CHEM, LIFE, MED, ENV, SOCH, SOCS, HUM, ART	Subject specialisation	A11
X12 text	University/ institution (name)	A12
X13 text	Total number of peer reviewed papers published <i>in most recent year</i>	A13
X14 number	Number of above as first author	A14
X15 number		A15

The X series questions were copied over from H2 so may not have been in schedule. Generally these can be completed from CV or whatever information was obtained at recruitment to project.

X01 ECR_ID is Country Code [=X07] + zero + 2 digits + Subject(=X11) + M/F[X03]. If ECR was in H1/H2 then you may use the same as H2.

X14/X15 Total number of peer reviewed papers published in most recent year. The original H2 question was for 2022, I changed that to the ambiguous 'most recent year' as I am not sure which year (2022 or 2023) we should be targeting. Either this

information comes direct from ECR or we use something like Google Scholar/Research Gate to get an independent measure.

Sections A-F are as in the final version of the interview schedule. [ECRs_Harbingers 3_Pilot Interview_schedule_1610DN.pdf¹]

During the course of H2 we were able to develop our analysis 'full-text' answers. For H3 we hope to further refine this technique. Thus for all questions we would like Quotes and Comments, the coded answer is of secondary importance. Coding is only requested for questions which have shown a high response rate when asked in previous Harbingers interviews.

Coding for sections A-F will be

a) Bimodal ,that is seeking **Yes** or **No** with optionally **NA** if not asked/answered and **Doubtful** if hesitant or undecided. Remember we would like quotes and commentary in addition to 'Yes/No' and there *must* be explanatory text for 'D'

b) Lickert Scale (0-9) where 0 is equivalent to NA and there is a general progression from least (1) to most intensive or positive (9). Note questions D11 and E08 are carried over from H2 and are scaled 0-5.

c) date YYYY-MM-00

d) number

Do not add or remove rows or columns from the coding sheet.

Do not use the code column for anything other than the permitted code, only one code, no punctuation

Note: There are spreadsheet (.xlsx) and document text (.docx) versions of the code-sheet, you can use whichever is most convenient. The .docx version will be converted to .xlsx as part of the pre-processing workflow, the .xlsx is then ingested into a SQL database. Any text formatting is likely to be lost during this process so do not use bold, italics, fonts, colours or complex paragraphing. Also please do not use all caps except for occasional emphasis nor strings of dashes, underlines, line-ends etc.

For each question in the H3 Pilot codesheet we will assess the equivalence with questions in H2-R3 We look for the closest match according to a classification in the table below.

- | | | | |
|---|----------------|---|--|
| a | Unchanged | Fully equivalent, minor typos | |
| b | minor revision | Reworded to remove ambiguity, coding revised. | Assume that problems that occasioned revision were resolved at coding stage for earlier version. |

1 From: <dave.nicholas@ciber-research.uk>
To: <anthony.watkinson@btinternet.com>, "Abrizah Abdullah" <abrizah@um.edu.my>, 'Blanca Rodríguez Bravo' <blanca.rodriguez@unileon.es>, "BOUKACEM CHERIFA" <cherifa.boukacem-zeghmouri@univ-lyon1.fr>, 'Marzena Świgoń' <marzena.swigon@uwm.edu.pl>, "Jorge Revez" <jrevez@campus.ul.pt>, '许洁' <xuj@whu.edu.cn>, <eherman@univ.haifa.ac.il>, "DJ Clark" <David.Clark@ciber-research.uk>

Subject: Interview schedule up and running

Date: Mon, 16 Oct 2023 08:15:15 +0100

- c Changed Original was off-target. Still fishing for the same response, may be possible to compare Q&C but not coding.
- d Not comparable but related. The original Q may have been split into a series of questions . May be possible to compare Q&Cs as a group of related responses.
- e New/Deleted No Equivalent.

The 'AI' set: These are all the questions concerning Artificial Intelligence in the Pilot interview schedule: B01–B07, C08, D04, D10, D14, D16, D17, E02–E05, E09, E13, E22, F03–F05.

AI questions B01, B03, B04, and B06. In addition to Quote and Comment I would like to try a coding experiment with these questions. The coding is not a question to be asked of the ECR but an assessment by the interviewer. So for B01: 0=none (nothing to say, no interest) 9 = AI researcher. B03: 1 = 'Played with ChatGPT — 9 = 'current research is AI project'. B04 1= inessential — 9 'AI is their project'. For values between 1 and 9 make a judgement by comparing to the other ECRs you have interviewed.

B06 Do they think their experience and opinion of 'AI' typical when compared to colleagues and friends?

An 0–9 scale could be too fine grained, but remember you are assessing your ECRs relatively as a group.

- 0 no answer / has nothing to say
- 1,2 resistant, sceptical
- 3,4 less interested, cautious
- 5 Typical, average
- 6,7 more interested,
- 8,9 enthusiast, early adopter

Note also that we are tending to assume experience and enthusiasm are correlated. As with all these questions the important thing is the quotes and commentary.

An example:

H3-Q	CODE	QUOTE	COMMENT
B06	4	My colleagues have similar experiences and opinions to mine, but my friends (non-researchers) end up using AI more nonchalantly and even see it as a business opportunity with little effort through the rapid generation of content that can be sold. It's also a means of leisure.	This ECR was a AI beginner and a pessimist one. His quotes are crystal clear.

H3-Q Permitted Codes H3-P1 Schedule Question

R3-Q R3-match

A01	text		A. Job				
		A01	Job title		B01	a	

H3-Q Permitted Codes		H3-P1 Schedule	Question	R3-Q	R3-match
A02	YYYY			B02	a
A03	text	A02	Start date (year)	B03	b
A04	text	A03	Previous job title	B04	a
A05	Y/N	A04	Previous location of job	B05	a
A06	Y/N	A05	Do they teach as part of their contract?	B06	a
		A06	Do they teach if it is not part of their contract		
			Research groups		
A07	Y/N	A07	Are they part of research group (s)? [formal or informal]	B08	a
A08	Y/N	A08	If yes, is it international in membership?	B09	a
			B. 'AI' general questions		
B01	text [+L0-9]	B01	What, if any, is their experience of, or at least encounters with 'AI' —not just in an academic or work context.		e
B02	text	B02	If so, what kind of 'AI' are we talking about?		e
B03	text [+L0-9]	B03	What is the extent of their use or engagement. For instance (to be used to jolt the memory, if necessary, after answering question):		e
B04	text [+L0-9]	B04	How do they view 'AI', for instance (to be used to jolt the memory, if necessary, after answering the question):		e
B05	text	B05	Are they concerned about the use of 'AI' in any way? If so, why? If their concern is about what might happen rather than something they have observed or experienced to-date:		e
B05a	text	a	Do you think they will come about (now, in a		e

H3-Q Permitted Codes H3-P1 Schedule Question R3-Q R3-match

			year or two, this decade, in your working lifetime)		
B05b	text	b	In general, would these speculative uses be a good or bad thing?		e
B06	text [+L0-9]	B06	Do they think their experience and opinion of 'AI' typical when compared to colleagues and friends?		e
B07	text	B07	Has AI helped in any way to reduce/manage teaching and administrative workload? If so, does that mean you can devote more time on research work?		e

C. C. Career aims and reputation

C01	Y/N	C01	Are they currently aiming for a permanent academic career in a university or similar research organisation?	C49	a
C02	text	C02	Assessment How does their institution, national panels, and/or funders assess them?	D01- D23	d
C03	text	C03	Reputation How would you judge your <i>success</i> as a researcher (and that of others)?		e
C04	text	C04	Do they consider download data, social media indicators/ alternative indicators (i.e., altmetrics) to have a <i>reputational</i> value for them or others?	D08	b
C05	text	C05	Achieving visibility for their research outputs is argued as being important in building research reputation: do they agree?	D10	b

H3-Q Permitted Codes	H3-P1 Schedule	Question	R3-Q	R3-match
C06	text	C06		
		If yes, how do they achieve maximum visibility for their research outputs? Do they you use any in this list?	D11	b
C06a	Y/N	a) Tweeting a link	D12	a
C06b	Y/N	b) Putting on a social scholarly platform;	D13	a
C06c	Y/N	c) Placing in a virtual poster where the article is summarised;	D15	a
C06d	Y/N	d) Giving a presentation at a conference, seminar or workshop (including virtual) where the article is referenced	D16	a
C06f	Y/N	f) Writing in a professional magazine/research newsletter	D18	a
C06g	Y/N	g) Blogging	D19	a
C06h	text	h) Other methods	D20	b
C07	text	C07		
		Argued there is a need to improve the ways in which scientific research output is evaluated by funding agencies and academic institutions by the considering openness and transparency factors, such as OA, open data and outreach. What are your views on such a policy?	D22	a
C08	text	C08		
		Does AI have any implications for research reputation? If so, what are they?		e

D. D. General communications practices

Information discovery and information usage

D01	text	D01	Where do they go to search for formal scholarly communications? List in order of importance.	E01	a
D02	text	D02	If the documents cannot be obtained easily (through their library/virtual network?) where do they go next? Listed in order of importance.	E02	a
D03	Y/N	D03	Do they use smartphones to search for/find formal scholarly information, such as full-text papers?	E07	a
D04	Y/N text	D04	Has their searching and discovery behaviour been impacted/changed in any way by 'AI'?		e
D05	text	D05	Sharing/connecting In what ways do they share/disseminate their ideas and their early stage/interim results?	E13	b
D06	Y/N text	D06	If not raised above, do they share their ideas and/or early stage/interim results over general social media channels, such as Instagram, Twitter, Facebook, TikTok and LinkedIn?	E14	b
D07	Y/N text	D07	If not raised above do they share their ideas and/or early stage/interim results through scholarly social	E17	b

H3-Q	Permitted Codes	H3-P1	Schedule Question	R3-Q	R3-match
D08	text		networks/social media, such as ResearchGate and Academia.edu?	E25	a
		D08	How do they currently go about <i>forming</i> ties with their fellow researchers?		
D09	text		How do they currently <i>maintain</i> these ties?	E26	a
		D09			
D10	text		Does AI have any implications for connecting and/or research sharing?		e
		D10			
			Information evaluation and trust		
D11	text		When they have searched and found an article on a topic important to their research what criterion persuades them to read it:	E29	b
		D11			
			[On a scale of 0 as no importance to 5 as very high importance ask interviewee to rate these characteristics]		
D11a	0-5		a) the name/reputation of the author	E30	a
D11b	0-5		b) the type of peer review process which the article has undergone	E31	b
D11c	0-5		c) the editor of the journal and members of the editorial board	E32	b
D11d	0-5		d) the name of the journal	E33	b
D11e	0-5		e) the name of the publisher	E34	b
D11f	0-5		f) abstract	E35	a
D11g	0-5		g) journal impact factor	E36	a
D12	text [L1-5]		To what extent do they feel that the peer review	E37	b
		D12			

H3-Q Permitted Codes H3-P1 Schedule Question

R3-Q R3-match

			system vouches for the quality and trustworthiness of formally published research?		
D13	text	D13	How do they decide how to trust informally disseminated evidence in their own specialisms?	E38	b
D14	text	D14	What would make them suspect that published material was possibly AI generated?		e
D15	text	D15	Are they aware of bad science/questionable/low grade practice being undertaken in their field and subsequently published? This is the place to prompt about grey publishers	D15	e
D16	text	D16	Do they believe that the AI-associated potential for rapid production of low-quality scientific articles brings about a decline in the overall quality of research output, indeed, facilitated the growth of predatory journals and papermills?		e
D17	text	D17	Do they think AI is raising any other issues of scholarly integrity and ethics? If so, what are they and what can be done about it? Might have been covered in C30 (reputation).		e
E01	Y/N	E01	E. Authorship, writing and publishing Does their research team/department/university have a clear authorship policy? If so, what is it?	F01	b

H3-Q Permitted Codes		H3-P1	Schedule Question	R3-Q	R3-match
E02	Y/N text	E02	Are these policies changing/being challenged because of 'AI' becoming (sort of) another author?		e
E03	Y/N	E03	Has 'AI' helped publishing productivity?		e
E04	Y/N	E04	Used 'AI' as a tool for summarising scientific articles/ extract key information from complex texts to facilitate doing a literature review;		e
E05	Y/N	E05	Used 'AI' to detect gaps in knowledge to locate a topic for new research and to construct hypotheses;		e
Publishing					
E06	Y/N text	E06	Do they see traditional journals, whether or not open access ones, as still the main way of making research available?	F10	b
E07	Y/N text	E07	Do they think more informal modes of communication (e.g., preprints, tweets, blogs) will play a larger role?	F11	b
E08	text	E08	When choosing a journal to submit their paper to which factors rate most highly:	F16	
<i>[Score on a scale 0 as no importance to 5 as high importance]</i>					
E08a	0-5		a) it is a high impact factor journal	F17	b
E08b	0-5		b) it has much prestige in the discipline	F18	b
E08c	0-5		c) appropriateness of the audience	F19	b
E08d	0-5		d) the speed from	F20	b

H3-Q Permitted Codes	H3-P1 Schedule	Question	R3-Q	R3-match
		submission to publication		
E08e	0-5	e) it is open access	F21	b
E08f	0-5	f) the geographical location/origins of journal/publisher	F22	b
E08g	0-5	g) where it is indexed (e.g., SCOPUS, Web of Science) [we are aware of overlap with a) above]	F25	b
E08h	0-5	h) high standards of peer review	F24	b
E09	Y/N text	E09 Will 'AI' change their relative ratings or introduce any new factors?		e
E10	Y/N text	E10 Peer review Are they involved in responding to criticisms of their/their groups' publications?	F26	b
E11	Y/N text	E11 Have they done peer review themselves?	F27	b
E12	Y/N text	E12 Do they feel that the peer review system needs improving in any way?	F29	b
E13	text	E13 What do you think an AI-based peer-review should be capable of doing, if it is to replace the current system?	F32	d/e
E14	Y/N text	E14 Open science and access publishing Does their research team/department/university have a policy regarding OA publishing? If yes, what is it?	F35	a
E15	Y/N text	E15 Does their research team/department/university have a policy on avoiding predatory and questionable journals? If	F37	b

H3-Q Permitted Codes	H3-P1 Schedule	Question	R3-Q	R3-match
E16	Y/N text	E16	yes, what is it? Can they/their group afford to publish in open access journals, which are entirely open access – so called gold journals? And in journals which are mostly not open access – so-called hybrid journals?	F38, F39 c
E17	Y/N text	E17	Is the final peer reviewed version of their article placed in a repository in their own institution and if so, why?	F41, F42 b
E18a	Y/N text	E18	Preprint servers Do they consider a preprint to be: a) an alternative to; a traditional publication? Whether yes or no, why?	F64 b
E18b	Y/N text		Do they consider a preprint to be: b) a replacement for a traditional publication? Whether yes or no, why?	F65 b
E19	text	E19	As stated earlier, making available research results quickly and openly can be at the expense of quality and reproducibility and it is argued that there is evidence for this in the number of retractions of preprints and final versions of papers.	F66, F67 b
E19a	Y/N text		a) Have any researchers you know retracted a paper?	F66 b
E19b	Y/N text		b) Have any editors or publishers you know retracted a paper?	F67 b
E20	Y/N text	E20	Outreach Are they	F68, F69 b

			expected/mandated by funders or government/institutional authorities to make their research comprehensible to researchers in fields other than their own, or to policy makers, industry, citizens etc? If yes, how do they go about it (platforms/media used) and which audiences are targeted?		
E21	Y/N text	E21	If not, do they do it nevertheless? If so, what methods are used and what audiences targeted?	F70	
E22	Text	E22	Is 'AI' capable of facilitating/enhancing their outreach activities? In what ways? Have they seen evidence of any increase in outreach in result?	F71	d/e
F. Transformations					
F01	text	F01	What form do they think a transformed scholarly communications system might take?	G05	a
F02	Y/N text	F02	Do they think that journals will still have a central role to play in ten years' time?	G07	a
F03	text	F03	What role do you think libraries will have for researchers in ten years' time as compared to their current role, especially in light of the growing utilization of 'AI'?	G08	b
F04	Y/N text	F04	Will 'AI' be a transformational force? If so, in what ways? What will be the advantages and disadvantages of the transformations that will		e

H3-Q Permitted Codes H3-P1 Schedule Question**R3-Q R3-match**

F05	Y/N text	F05	take place? Will the use of 'AI' exacerbate existing disparities and inequalities, with people with access to AI-based tools speeding up their publication processes?		e
F06	text	F06	Finally, how are they faring now that the pandemic is over? This is a question, if relevant, about similarities/differences in ECRs' circumstances, attitudes and practices, compared to Covid-times. Any other areas of scholarly communication touched by AI but not picked up on previously?	H01	d
G00	text				e