

Assessing the Opportunities and Challenges of Artificial Intelligence Tools in Scientific Research: Evidence from Samangan University

Abdul Wali Sirat^{1*}, Musawer Hakimi², Baryali Sazish²

¹Department of History, Education Faculty, Samangan University, Samangan, Afghanistan

²Department of Computer Science, Education Faculty, Samangan University, Samangan, Afghanistan

*Corresponding author: abdulwalis123@gmail.com

Abstract. Artificial intelligence (AI) tools are changing how research is designed, analysed and written, yet little is known about how prepared universities in Afghanistan are to use them. This study examines perceived opportunities, challenges, infrastructure, data quality and privacy concerns around AI tools among 300 students, lecturers and researchers at Samangan University, and how these perceptions relate to readiness for big-data analytics. A cross-sectional self-report survey used a five-point Likert instrument covering seven constructs; 74.7% of respondents were students. Data were analysed with descriptive statistics, group comparisons, correlations and multiple regression. Perceptions sat close to the scale midpoint on every construct (means 2.96–3.06), and 75.3% of respondents fell in the moderate band of a descriptive readiness index. Respondents who had attended AI or data-analysis training reported higher statistical competence (3.12 vs 2.86, $p = .003$); overall readiness did not differ significantly ($p = .057$). These comparisons are observational and unadjusted. In the regression model, perceived opportunities, statistical competence, infrastructure and data quality were positively associated with perceived analytical effectiveness, while perceived challenges were negatively associated with it ($R^2 = .455$). The results describe perceptions, not objective performance or causal effects, and should not be generalised beyond this student-dominated, single-institution sample.

Keywords: Artificial Intelligence, Big Data Analytics, Higher Education, Research Readiness, Samangan University

1 Introduction

Over the last few years, artificial intelligence (AI) tools have moved out of specialist laboratories and into everyday academic work. Researchers now use them to search the literature, clean and analyse data, draft text and even suggest hypotheses [1], [2]. Large language models in particular have put these capabilities within reach of anyone with an internet connection, and that has set off a lively debate about quality, authorship and trust in scientific work [3], [4].

The opportunities are easy to list. AI tools can shorten the time spent on routine tasks, support writers who work in a second language, and let small teams handle datasets that once needed a large group [2], [5]. Reviews of research in higher education describe similar expectations, together with growing use by both students and academic staff [6], [7]. The risks are just as well documented. Language models can produce fluent text that is simply wrong [8], they can reproduce biases from their training data [9], and they raise uncomfortable questions about academic integrity and about what researchers really understand when a tool does the analysis for them [10], [11]. Concerns about privacy and responsible data handling run through most national and institutional ethics guidelines [12], [13].

Most of this evidence comes from well-resourced systems. Evidence from universities where connectivity is unstable, software licences are expensive and formal training is limited is much thinner. In those settings, whether AI tools help or hinder research probably depends less on the tools than on whether people have the skills, the data and the infrastructure to use them well. Students' own accounts point in the same direction: they see benefits, but they also worry about over-reliance and accuracy [14]. Technology acceptance research has long argued that perceived usefulness and ease of use shape adoption [15], and that organisational support decides whether good intentions turn into actual use [16].

Evidence on AI readiness in universities usually comes from self-report questionnaires. Such instruments capture what respondents believe about their own skills and about AI tools. That is informative about attitudes and perceived readiness, but it is not a direct measure of analytical accuracy, research productivity or the quality of scientific output, and the present study should be read in that light.

This study looks at Samangan University in northern Afghanistan, where, to our knowledge, little published evidence exists on how members of the university view AI tools in research. The practical problem is that decisions about training and investment are being made without local data on perceived opportunities, challenges and readiness. Three questions guide the work. First, how do students, lecturers and researchers perceive the opportunities and challenges of AI tools in research, and how ready are they for big-data analytics? Second, how does readiness vary by age, role, education, experience or training? All respondents in this sample were male, so gender differences could not be assessed. Third, which factors are associated with perceived analytical effectiveness when AI tools are used? Drawing on the literature above, we expected opportunity perceptions, statistical competence, infrastructure and data quality to be positively associated with perceived analytical effectiveness, and perceived challenges and privacy concern to be negatively associated with it. The purpose of the article is to describe these patterns of perception and to give university leaders perception-based evidence they can consider when planning training and infrastructure.

2 Methods

2.1 Design and participants

We used a cross-sectional quantitative survey. The final sample contained 300 respondents from Samangan University: 224 students, 51 lecturers and 25 researchers or other research staff. The available manuscript does not document the target-population frame, eligibility criteria, recruitment method, sampling procedure, survey dates or administration mode, nor the numbers invited, responding or excluded. These details cannot be reconstructed reliably from the final sample size alone and should be reported from the original fieldwork records if available. The manuscript also contains no record of the consent procedure, confidentiality arrangements, institutional authorisation or ethics approval. Accordingly, no approval number or undocumented procedure is asserted here; the authors should verify these details against the original records before submission.

Students made up 74.7% of the respondents, so the findings mainly reflect student perceptions. The questionnaire recorded respondents' research experience and prior big-data experience (Section 2.2), but the analyses reported here do not condition on these characteristics or on whether students had completed a research-methods course, a thesis or a research project, or had used AI tools in research. The manuscript does not provide the operational eligibility criteria used for student participants or the frequency distribution of the research-experience variable by respondent role. Although research experience is listed as a questionnaire variable, its observed values and coding are not supplied in the materials available for this revision. These distributions therefore cannot be reported without the original respondent-level data. The analyses remain unadjusted for research experience, completion of research-methods courses or projects, and prior use of AI tools in research. No claim is made that the sample is statistically representative of the university's research community, and student perceptions are not treated as evidence of institutional research readiness.

2.2 Instrument and scoring

The questionnaire had two parts. The first collected background information: gender, age, role, education level, research experience, prior experience with big data, and whether the respondent had attended any training in AI or data analysis. The second part contained 35 statements rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). These statements measured seven constructs: statistical competence (5 items), perceived opportunities of AI tools (6 items), perceived challenges (6 items), infrastructure support (5 items), data quality (4 items), privacy and security concern (4 items) and perceived analytical effectiveness when using AI tools (5 items), a self-reported judgement of how effectively respondents can analyse data with AI tools. Each construct score is the mean of its items, so it ranges from 1 to 5. A higher score means more of the construct; for challenges and privacy, a higher score therefore means greater concern.

Table A1 (Appendix A) lists the constructs, item counts and scoring direction. Perceived analytical effectiveness is a self-report measure; it does not directly measure analytical accuracy, research productivity, task performance or the quality of scientific output. Cronbach's alpha is an index of internal consistency and does not, by itself, establish construct validity [17]. The materials available for this revision do not identify whether the questionnaire items were newly written or adapted, provide the original item wordings or source instruments, or document expert review, translation, pilot testing, revisions, reliability coefficients calculated from item-level responses, or factor analysis. The original 35 item wordings and instrument-development records must be checked and supplied by the authors. Because these records and original item-level data were not provided, construct-specific Cronbach's alpha values and factor-analytic results cannot be calculated or claimed here. The seven constructs are therefore treated as theoretically defined scales, without a claim that their empirical distinctness has been established. Until such evidence is reported, the seven constructs are treated as theoretically defined scales and are not claimed to be empirically distinct.

2.3 Readiness index and cut-offs

Big-data analytics readiness was summarised as a composite index calculated from the construct scores. Respondents were placed in three bands: low (index up to 2.33), moderate (2.34 to 3.67) and high (above 3.67). The index is treated here as a descriptive composite, not as a validated scale. The cut-offs coincide with dividing the 1–5 range into three equal intervals ($1 + 4/3 = 2.33$; $1 + 8/3 = 3.67$). The bands are therefore descriptive, have not been validated against AI skills, tool use or research performance, and are not diagnostic or institutional readiness classifications; group comparisons of the index use its continuous scores (Table 3). Challenges and privacy/security concern are barriers rather than enablers, so how they enter the index must be stated explicitly. The exact readiness-index formula used for the reported results is not documented in the supplied manuscript. In particular, the available materials do not establish which constructs were included, the weighting and averaging procedure, the treatment of missing responses, or whether challenges and privacy/security concern were reverse-scored. The cut-offs correspond to three equal intervals of a 1–5 scale ($1 + 4/3 = 2.33$; $1 + 8/3 = 3.67$), but the materials do not establish whether these cut-offs were fixed before analysis. Until the original analysis syntax or calculation records are checked, the reported readiness-index results should be regarded as provisional; the index and all results based on it must be recalculated from the original data if the implemented formula cannot be verified.

2.4 Data analysis

Analyses were run in Python 3 (pandas, SciPy and Matplotlib). We first described the sample and each construct with frequencies, means and standard deviations. Group differences in readiness were tested with Welch's *t*-test (two groups) and one-way ANOVA (three or more groups); the link between training and readiness band was

tested with a chi-square test. Associations among constructs were examined with Pearson correlations. Finally, an ordinary least squares regression took perceived analytical effectiveness as the dependent variable and the other six constructs as predictors; variance inflation factors (VIF) were checked for multicollinearity. Linearity, residual-versus-fitted patterns, residual normality, independence of errors and influential observations are not reported in this version. Analyses were reported as having been run in Python 3 using pandas, SciPy and Matplotlib. Descriptive statistics, Welch's t-tests, one-way ANOVA, a chi-square test, Pearson correlations and an ordinary least squares regression were used as described above. The manuscript reports VIF values of 1.23–1.39 and tolerance values of 0.72–0.81. However, the supplied materials do not include the original respondent-level dataset, analysis script, residual plots, studentised residuals, Cook's distances, leverage values, heteroscedasticity test or robust-standard-error output. These diagnostics and confidence intervals cannot be truthfully reported as completed without rerunning the analyses on the original data. The reported model summary is $R^2 = .455$, adjusted $R^2 = .444$, $F(6, 293) = 40.80$, $p < .001$, with $N = 300$; it should be verified against the original output before publication. The significance level was set at .05 throughout, without adjustment for the number of comparisons, so the subgroup results are exploratory.

Two design features limit what the analyses can show. First, the comparison by training status is observational: respondents were not assigned to training, and no covariate-adjusted estimate is reported here. Second, all constructs were measured with one self-report questionnaire completed by the same respondents at one time, so common-method variance and related response biases are possible [18]; no procedural or statistical remedy was applied, and a single-factor test, were it used, would not prove their absence. The readiness index is not entered in the regression, and correlations that involve the index are not interpreted as independent evidence because the index is built from the construct scores. The available materials do not establish whether the original item-level dataset contains complete usable measures for age, role, education, research experience, prior big-data experience and training status, or whether a covariate-adjusted training comparison was computed. No adjusted model is therefore added. The training comparisons are interpreted as unadjusted and observational; any adjusted model, confidence intervals and diagnostics should be produced only after confirming the relevant variables and rerunning the analysis on the original data.

3 Results

3.1 Sample characteristics

Table 1 and Fig. 1 describe the 300 respondents. Men made up 64.0% of the sample and women 36.0%. The mean age was 34.5 years ($SD = 9.8$, range 18–50). Students were the largest group (74.7%), followed by lecturers (17.0%) and researchers or research staff (8.3%). About four in ten respondents (42.0%) said they had attended some form of AI or data-analysis training. Because three quarters of respondents are students,

overall means are dominated by student responses; role-specific readiness scores are given in Table 3.

Table 1. Characteristics of the respondents (*N* = 300)

Characteristic	Category	<i>n</i>	%
Gender	Male	300	100.0
Age group (years)	18–24	55	18.3
	25–34	93	31.0
	35–44	90	30.0
	45–50	62	20.7
Role	Student	224	74.7
	Lecturer	51	17.0
	Researcher / staff	25	8.3
Education level (coded 1–4)	1	97	32.3
	2	134	44.7
	3	45	15.0
	4	24	8.0
Prior big-data experience (coded 1–4)	1	95	31.7
	2	100	33.3
	3	78	26.0
	4	27	9.0
AI / data-analysis training	Attended	126	42.0
	Not attended	174	58.0

Note. Education level and prior big-data experience are shown with their original numeric codes. Education level and prior big-data experience are shown using their original numeric codes because the category labels and codebook were not included in the materials available for this revision. The authors should insert the exact labels from the original questionnaire/codebook and verify that the same coding was used in the ANOVAs before submission; labels have not been inferred or invented.

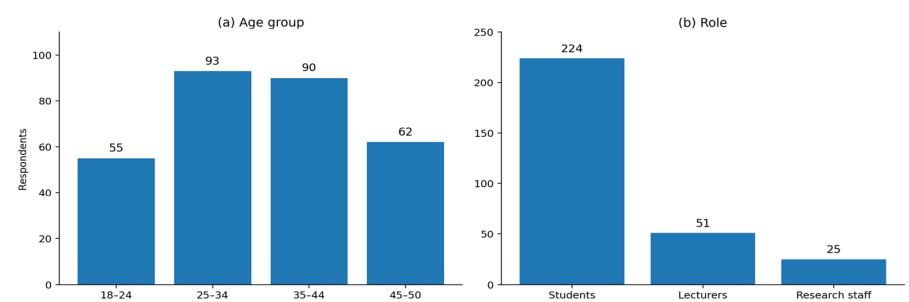


Fig. 1. Respondents by age group and role.

3.2 Perceptions across the seven constructs

Table 2 and Fig. 2 show a strikingly even picture. Every construct had a mean close to the midpoint of the scale, from 2.96 for data quality to 3.06 for perceived opportunities. Perceived challenges ($M = 3.01$) were rated about as high as perceived opportunities, which suggests that respondents see both sides of AI at the same time rather than leaning clearly toward enthusiasm or alarm. The readiness index averaged 2.98 ($SD = 0.56$) and ranged from 1.42 to 4.69.

Table 2. Descriptive statistics for the study constructs (scale 1–5)

Construct	Items	<i>M</i>	<i>SD</i>	Min	Max
Statistical competence	5	2.97	0.78	1.00	5.00
Perceived opportunities	6	3.06	0.71	1.17	4.83
Perceived challenges	6	3.01	0.75	1.00	5.00
Infrastructure support	5	2.99	0.74	1.00	5.00
Data quality	4	2.96	0.78	1.00	5.00
Privacy/security concern	4	2.98	0.78	1.00	5.00
Perceived analytical effectiveness	5	3.02	0.73	1.20	5.00
Big-data readiness index (descriptive composite)	–	2.98	0.56	1.42	4.69

Note. For challenges and privacy/security, higher scores indicate greater concern.

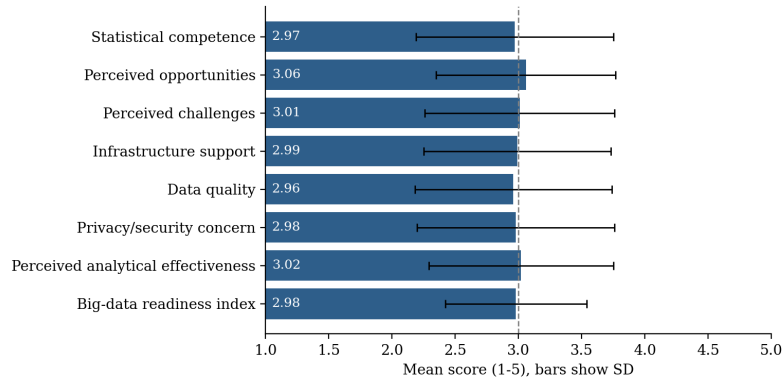


Fig. 2. Mean scores (± 1 SD) for each construct. The dashed line marks the scale midpoint. The figure was re-plotted from the values in Table 2 to carry the label ‘perceived analytical effectiveness’.

3.3 Readiness levels

Most respondents fell in the middle band (Fig. 3a): 226 (75.3%) were classed as moderately ready, 38 (12.7%) as low and 36 (12.0%) as high. These bands are descriptive

(Section 2.3). Respondents who had attended training were more often in the high band and less often in the low band (Fig. 3b). Among those who had attended training, 16.7% were in the high band and 7.1% in the low band; among those who had not, the figures were 8.6% and 16.7%. This pattern was statistically significant, $\chi^2(2) = 9.20$, $p = .010$, Cramér's $V = .18$, although the effect is small.

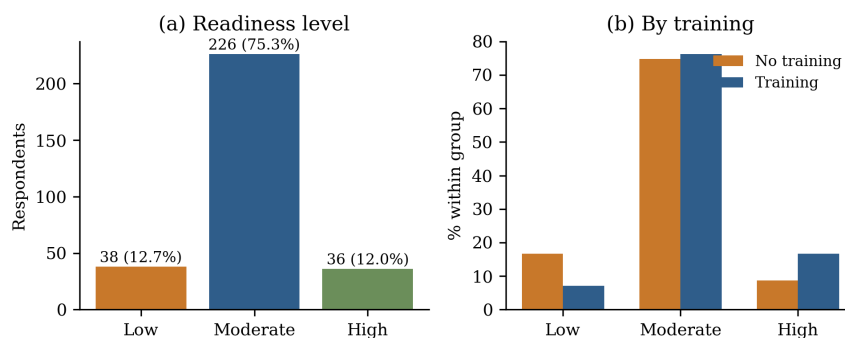


Fig. 3. Distribution of readiness levels (a) and readiness level by training status (b).

3.4 Differences between groups

Table 3 and Fig. 4 compare the readiness index across respondent characteristics. Because all respondents were male, gender differences could not be examined. Readiness did not differ significantly by role or education level, and it was unrelated to age ($r = .06$, $p = .317$). Training showed a modest advantage that fell just short of the conventional threshold (3.06 vs 2.93; Cohen's $d = 0.22$). The more telling differences appeared in the separate constructs. Respondents who had attended training reported higher statistical competence (3.12 vs 2.86, $t(280.6) = 2.97$, $p = .003$) and slightly higher perceived opportunities (3.16 vs 2.99, $t(276.5) = 2.02$, $p = .044$), but did not differ in perceived challenges ($p = .524$). These training comparisons are observational. Respondents were not assigned to training, and those who attended may already have differed in confidence, experience or motivation (self-selection), so the differences are descriptive and are not estimates of the effect of training. Statistical competence also rose with prior big-data experience, from 2.81 at the lowest level to 3.43 at the highest, $F(3, 296) = 4.69$, $p = .003$. Perceived opportunities varied by role, with researchers highest (3.27), students in the middle (3.08) and lecturers lowest (2.88), but this difference was borderline, $F(2, 297) = 2.98$, $p = .052$, and the researcher group contains only 25 people.

Table 3. Big-data readiness index by respondent characteristics

Factor	Group	n	M	SD	Test
Training	No training	174	2.93	0.57	$t(275.1) = 1.91, p = .057$
	Training	126	3.06	0.55	
Role	Student	224	2.97	0.57	$F(2, 297) = 1.00, p = .371$
	Lecturer	51	2.99	0.52	
	Researcher / staff	25	3.13	0.56	
Education level	1	97	2.99	0.59	$F(3, 296) = 0.08, p = .972$
	2	134	2.97	0.54	
	3	45	3.01	0.59	
	4	24	3.00	0.54	
Prior big-data experience	1	95	2.94	0.55	$F(3, 296) = 2.11, p = .099$
	2	100	2.95	0.62	
	3	78	3.00	0.46	
	4	27	3.23	0.57	

Note. Welch's *t*-test was used for two groups and one-way ANOVA for three or more groups. Degrees of freedom for Welch tests are approximate. The index is a descriptive composite; education level and big-data experience are shown with their original numeric codes (see the note to Table 1). Comparisons are unadjusted.

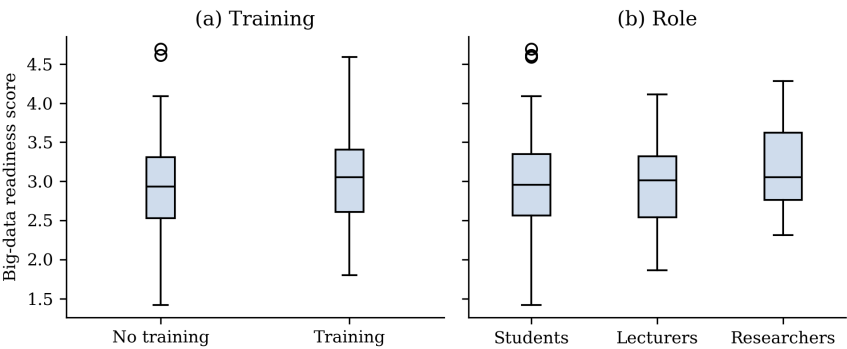


Fig. 4. Readiness index by training status (a) and by role (b). Boxes show the interquartile range and the line marks the median.

3.5 Relationships among constructs

Table 4 and Fig. 5 present the correlations. Statistical competence and perceived opportunities were moderately related ($r = .46$), and both were positively related to infra-

structure support and data quality. Perceived challenges moved in the opposite direction: they correlated negatively with infrastructure ($r = -.31$), data quality ($r = -.29$) and perceived analytical effectiveness ($r = -.27$), and positively with privacy concern ($r = .43$). Perceived analytical effectiveness was most strongly related to perceived opportunities ($r = .57$) and statistical competence ($r = .50$). The readiness index correlated strongly with several constructs, but because it is built from them, these coefficients partly reflect how the index was constructed and should not be read as independent findings. This applies in particular to its correlation with perceived analytical effectiveness ($r = .77$), which would be mathematically coupled if that construct is among the index components. The correlations among the seven constructs come from one self-report questionnaire and may be inflated by shared method variance [18].

Table 4. Pearson correlations among the study constructs ($N = 300$)

Construct	1	2	3	4	5	6	7
1. Stat. competence	–						
2. Opportunities	.46**	–					
3. Challenges	-.12*	-.17**	–				
4. Infrastructure	.35**	.36**	-.31**	–			
5. Data quality	.32**	.31**	-.29**	.41**	–		
6. Privacy concern	-.09	-.11	.43**	-.19**	-.16**	–	
7. Perceived analytical eff.	.50**	.57**	-.27**	.42**	.39**	-.15*	–
8. Readiness index (composite; coupled with 1–7)	.74**	.57**	-.33**	.73**	.72**	-.20**	.77**

Note. One asterisk denotes $p < .05$ and two asterisks denote $p < .01$.

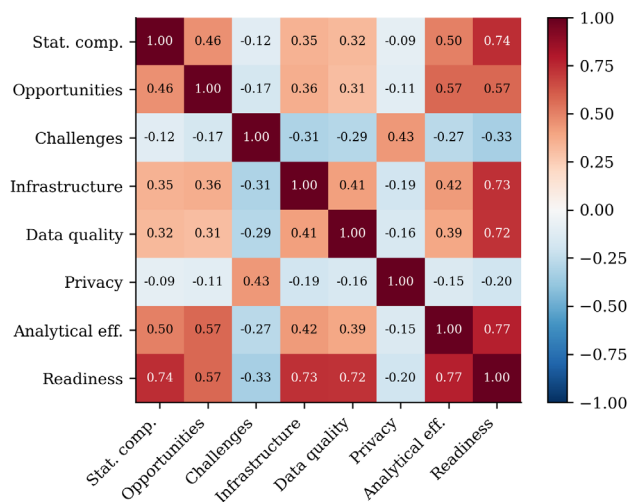


Fig. 5. Correlation matrix of the study constructs. Where the figure labels a row or column ‘analytical effectiveness’, it refers to perceived analytical effectiveness.

3.6 Predictors of perceived analytical effectiveness

The regression model was significant and accounted for 45.5% of the variance in perceived analytical effectiveness (adjusted $R^2 = .444$), $F(6, 293) = 40.80$, $p < .001$ (Table 5, Fig. 6). VIF values ranged from 1.23 to 1.39 (tolerance 0.72 to 0.81), so multicollinearity was not a concern. Perceived opportunities had the largest standardised coefficient ($\beta = .36$), followed by statistical competence ($\beta = .25$). Infrastructure support ($\beta = .13$) and data quality ($\beta = .11$) each made a smaller but significant contribution, and perceived challenges were negatively associated with perceived effectiveness ($\beta = -.11$). Privacy and security concern showed no independent association ($\beta = .01$, $p = .890$). These are associations among self-reported perceptions from one cross-sectional questionnaire; they are not causal effects, and the model does not include the readiness index. Residual and influence diagnostics beyond VIF are not reported (Section 2.4).

Table 5. Multiple regression predicting perceived analytical effectiveness

Predictor	B	SE	β	t	p
(Constant)	.80	.28	–	2.89	.004
Statistical competence	.23	.05	.25	4.88	< .001
Perceived opportunities	.37	.05	.36	7.11	< .001
Perceived challenges	-.11	.05	-.11	-2.18	.030
Infrastructure support	.13	.05	.13	2.53	.012
Data quality	.11	.05	.11	2.28	.023
Privacy/security concern	.01	.05	.01	0.14	.890

Note. $R^2 = .455$; adjusted $R^2 = .444$; $F(6, 293) = 40.80$, $p < .001$. B = unstandardised coefficient; β = standardised coefficient.

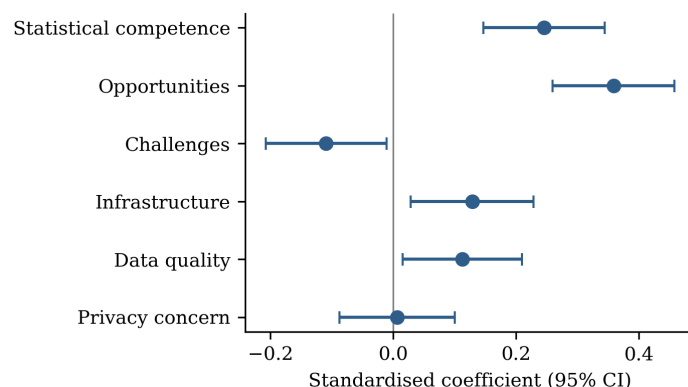


Fig. 6. Standardised regression coefficients with 95% confidence intervals. Any outcome label in this figure refers to perceived analytical effectiveness.

4 Discussion

The most noticeable result is how evenly respondents sit around the middle of the scale. They neither embrace AI tools in research nor reject them, and they rate the benefits and the difficulties about equally. That fits a community in the early stages of adoption, where experience is uneven and opinions are still forming, and it echoes the mixed picture painted by wider discussions of generative AI in research [19]. It may also partly reflect a tendency to choose central answers on Likert scales, so we would not push the interpretation too far. Because three quarters of respondents are students, this picture mainly describes student perceptions; it is not evidence of the research readiness of the university as an institution.

Training stands out as the one background factor associated with higher self-reported statistical competence. Trained respondents rated their statistical competence higher and were about twice as likely to be in the high readiness band, which matches the view that AI use pays off mostly for people who already have the skills to judge its output [5], [11]. We should be careful, however. The design is cross-sectional, so we cannot say that training produced the difference; people who sign up for training may have been more confident or more experienced to begin with. The training effect on the overall readiness index was small and just missed significance, so the evidence is suggestive rather than conclusive. The comparison is also unadjusted for age, role, education or prior experience. Methods such as multivariable adjustment or propensity-score analysis [20] could be applied to the item-level data, provided the groups overlap enough and the assumptions can be checked, but they would still address only observed covariates and could not rule out self-selection on unmeasured traits.

Age, education and role were not clearly related to the readiness index in this sample. Because all respondents were male, the study cannot estimate or compare gender differences. This should be read cautiously: the researcher group is small, the study was not designed to detect small differences, and the data cannot show whether access and skills, rather than demographic background, are the levers for change. The slightly higher opportunity ratings among researchers and the lower ones among lecturers are worth watching, but the researcher group is small and the difference was borderline.

The regression results line up with technology acceptance research. Perceiving AI tools as useful was the strongest correlate of perceived effective use [15], and the contribution of infrastructure support is consistent with the argument that organisational conditions shape whether intentions turn into practice [16]. The independent role of data quality fits the emphasis on well-managed, findable and reusable data in current research data principles [21]. The weak and non-significant role of privacy concern deserves a cautious reading. It is strongly correlated with perceived challenges ($r = .43$), so its effect may simply be absorbed by that broader construct; it is also possible that respondents with limited hands-on experience have not yet run into the practical consequences of privacy problems. Our data cannot separate these explanations. The model explains 45.5% of the variance in the outcome, which leaves most of the variation unexplained by these constructs, and the coefficients are statistically significant without necessarily being practically large.

The outcome is perceived analytical effectiveness, not measured performance. Respondents' ratings do not show that AI tools raised analytical accuracy, productivity or the quality of scientific output, and the conclusions are restricted to perceptions. The associations may also be inflated by shared method variance, since predictors and outcome come from the same respondents in one questionnaire [18], and the correlations that involve the readiness index are partly built into its construction. Internal-consistency and factor-analytic evidence for the seven constructs is not yet reported (Section 2.2).

Several practical implications follow, with the caution that none was tested here. Universities that want to benefit from AI tools in research may wish to consider short, hands-on training in statistics and data handling and steady investment in connectivity and data management, but these data do not show that either would change research practice or outcomes. Policy matters too. Guidance on acceptable AI use in research writing is needed [22], [23], and institutions should be careful about relying on automated AI-text detectors to enforce it, because such detectors have been shown to misclassify the writing of non-native English speakers [24].

The study has limitations. It covers one institution, relies on self-reported perceptions rather than observed behaviour, and is cross-sectional. The readiness index overlaps with several of the constructs it is compared with, the researcher subgroup is small ($n = 25$), and the analysis used construct scores rather than item-level models. In addition, 74.7% of respondents were students; the sampling procedure, response rate and eligibility criteria are not documented in this version; the questionnaire has no reported validation evidence; the readiness categories are descriptive; training status was self-selected and unadjusted; common-method variance was not assessed; residual diagnostics beyond VIF are not reported; and many subgroup tests were run at the .05 level

without adjustment. Future work could follow respondents over time, add interviews to explain the midpoint pattern, compare several Afghan universities, validate the instrument and the index with item-level data, include objective measures of analytical performance, and test whether training changes research practice.

5 Conclusions and Suggestions

Members of Samangan University who responded to the survey see both promise and difficulty in AI tools for research, and most fell in the moderate band of a descriptive readiness index. Readiness did not differ clearly by age, role or education; because all respondents were male, gender differences could not be assessed, but training and prior big-data experience were associated with higher statistical competence, and perceived opportunities, statistical competence, infrastructure and data quality were each associated with higher perceived analytical effectiveness, while perceived challenges were associated with lower perceived effectiveness. These findings answer the three research questions for this sample: perceptions are balanced, background characteristics showed little relation to readiness, and skills, infrastructure and a favourable view of AI go together with perceived effective use. Because the sample is dominated by students, the design is cross-sectional and all measures are self-reported, these findings describe perceptions and associations; they do not show objective performance, institutional readiness or causal effects.

As options consistent with these perceptions, rather than recommendations proven by the data, we suggest that the university consider (1) offering practical training in statistics and AI-assisted analysis to students and staff, (2) investing in reliable connectivity and basic research data management, and (3) publishing clear guidance on responsible AI use. Researchers should test these suggestions with longitudinal or interventional designs and objective performance measures, since the present data cannot show cause and effect.

Appendix A. Measurement structure

Table A1 lists the seven constructs, the number of items and the scoring direction reported in this manuscript. Item codes, wordings and sources are not available in the manuscript and must be supplied by the authors from the original questionnaire.

Table A1. Constructs and items of the questionnaire (35 items, five-point Likert scale)

Construct	Items	Scoring direction	Item codes and wording	Source / adaptation
Statistical competence	5	Higher = more competence	[to be supplied]	[to be supplied]
Perceived opportunities	6	Higher = more perceived opportunity	[to be supplied]	[to be supplied]
Perceived challenges	6	Higher = greater concern (barrier)	[to be supplied]	[to be supplied]
Infrastructure support	5	Higher = more support	[to be supplied]	[to be supplied]
Data quality	4	Higher = better data quality	[to be supplied]	[to be supplied]
Privacy and security concern	4	Higher = greater concern (barrier)	[to be supplied]	[to be supplied]
Perceived analytical effectiveness	5	Higher = higher self-rated effectiveness	[to be supplied]	[to be supplied]

Note. Scoring direction for ‘data quality’ and ‘infrastructure support’ is inferred from the construct names and the Methods text (‘a higher score means more of the construct’); the authors should confirm it against the item wording.

Acknowledgement

The authors thank the students, lecturers and researchers at Samangan University who took the time to complete the survey.

References

- [1] H. Wang *et al.*, “Scientific discovery in the age of artificial intelligence,” *Nature*, vol. 620, pp. 47–60, 2023, doi: 10.1038/s41586-023-06221-2.
- [2] M. Salvagno, F. S. Taccone, and A. G. Gerli, “Can artificial intelligence help for scientific writing?,” *Crit. Care*, vol. 27, Art. no. 75, 2023, doi: 10.1186/s13054-023-04380-2.
- [3] E. A. M. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, “ChatGPT: Five priorities for research,” *Nature*, vol. 614, pp. 224–226, 2023, doi: 10.1038/d41586-023-00288-7.
- [4] H. H. Thorp, “ChatGPT is fun, but not an author,” *Science*, vol. 379, no. 6630, p. 313, 2023, doi: 10.1126/science.adg7879.
- [5] E. Kasneci *et al.*, “ChatGPT for good? On opportunities and challenges of large language models for education,” *Learn. Individ. Differ.*, vol. 103, Art. no. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

- [6] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," *Int. J. Educ. Technol. High. Educ.*, vol. 16, Art. no. 39, 2019, doi: 10.1186/s41239-019-0171-0.
- [7] H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," *Int. J. Educ. Technol. High. Educ.*, vol. 20, Art. no. 22, 2023, doi: 10.1186/s41239-023-00392-8.
- [8] Z. Ji *et al.*, "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, Art. no. 248, 2023, doi: 10.1145/3571730.
- [9] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in *Proc. ACM Conf. Fairness, Accountability, and Transparency (FAccT)*, 2021, pp. 610–623, doi: 10.1145/3442188.3445922.
- [10] D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innov. Educ. Teach. Int.*, vol. 61, no. 2, pp. 228–239, 2024, doi: 10.1080/14703297.2023.2190148.
- [11] L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," *Nature*, vol. 627, pp. 49–58, 2024, doi: 10.1038/s41586-024-07146-0.
- [12] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, pp. 389–399, 2019, doi: 10.1038/s42256-019-0088-2.
- [13] W. Holmes *et al.*, "Ethics of AI in education: Towards a community-wide framework," *Int. J. Artif. Intell. Educ.*, vol. 32, pp. 504–526, 2022, doi: 10.1007/s40593-021-00239-1.
- [14] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 20, Art. no. 43, 2023, doi: 10.1186/s41239-023-00411-8.
- [15] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, 1989.
- [16] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003.
- [17] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Educ.*, vol. 2, pp. 53–55, 2011, doi: 10.5116/ijme.4dfb.8dfd.
- [18] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: A critical review of the literature and recommended remedies," *J. Appl. Psychol.*, vol. 88, no. 5, pp. 879–903, 2003, doi: 10.1037/0021-9010.88.5.879.
- [19] Y. K. Dwivedi *et al.*, "Opinion paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, Art. no. 102642, 2023, doi: 10.1016/j.jinfomgt.2023.102642.
- [20] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.
- [21] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, Art. no. 160018, 2016, doi: 10.1038/sdata.2016.18.
- [22] B. D. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang, "ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language model in scholarly publishing," *J. Assoc. Inf. Sci. Technol.*, vol. 74, no. 5, pp. 570–581, 2023, doi: 10.1002/asi.24750.

- [23] S. A. Bin-Nashwan, M. Sadallah, and M. Bouteraa, "Use of ChatGPT in academia: Academic integrity hangs in the balance," *Technol. Soc.*, vol. 75, Art. no. 102370, 2023, doi: 10.1016/j.techsoc.2023.102370.
- [24] W. Liang, M. Yuksekgonul, Y. Mao, E. Wu, and J. Zou, "GPT detectors are biased against non-native English writers," *Patterns*, vol. 4, no. 7, Art. no. 100779, 2023, doi: 10.1016/j.patter.2023.100779.