

## Создание и апробация опросника UX встроенных ИИ-помощников

Научный руководитель – Горбунова Елена Сергеевна

*Коваленко Ю.Д.<sup>1</sup>, Зайцева Д.Р.<sup>2</sup>*

1 - Национальный исследовательский университет «Высшая школа экономики», Факультет социальных наук, Москва, Россия, *E-mail: idkovalenko@edu.hse.ru*; 2 - Национальный исследовательский университет «Высшая школа экономики», Факультет социальных наук, Москва, Россия, *E-mail: drzaytseva@edu.hse.ru*

В настоящее время растет запрос на использование ИИ-инструментов при решении задач - это позволяет уменьшить затраты человеческих ресурсов, оптимизировать организацию деятельности и повысить скорость и в некоторых случаях качество решения [6, 10]. Однако на данный момент нет инструмента, который бы должным образом отражал восприятие инструмента пользователем при взаимодействии с ним. Одними из основных конструкторов, которые выделяют в статьях и которые относятся к UX ИИ-помощника, являются персонализация функций ИИ-инструмента, доверие к нему и удобство использования. И существующие инструменты, пересекающиеся по теме, не покрывают все аспекты встроенного ИИ-помощника (чтобы это компенсировать, нужно использовать несколько шкал в комплексе), а также не все из них адаптированы на русский язык.

Наша исследовательская работа направлена на создание инструмента, позволяющего полноценно отследить аспекты взаимодействия пользователей с опросником и их удовлетворенность. Актуальность создания данного инструмента основывается на повышенном спросе на использование ИИ-инструментов для облегчения решения задач и, соответственно, необходимости учитывания запросов пользователей при разработке инструментов. Так, наша целевая аудитория - это разработчики ИИ-инструментов, а именно компании, внедряющие ИИ в свои приложения и сервисы.

В процессе создания нашего инструмента, мы опирались на следующие определения:

- UX (user experience, пользовательский опыт) - впечатления пользователя, возникающие в результате использования и/или предстоящего использования продукции, системы или услуги [15].
- ИИ-помощник, или диалоговый агент - компьютерная программа, которая поддерживает диалог с пользователем, используя естественный язык, и динамически генерирует ответы с учетом контекста. Источники: [3, 11, 13] Allouch et al., 2021; Poser et al., 2022; Radziwill & Benton, 2017

Субконструктами в нашем опроснике являются:

- 1) Доверие пользователя к ИИ-инструменту - установка, что агент поможет человеку достичь целей в ситуации неопределенности и уязвимости. [9]
- 2) Персонализация - уровень адаптации ответов системы с целью удовлетворения потребностей пользователя [7]
- 3) Удобство использования - часть пользовательского опыта, предполагающая оценку эффективности и простоты использования инструмента.

Субконструкты выделены на основе теоретических концептуализаций других инструментов, которые подходят для ситуации использования встроенного ИИ-помощника, а именно: TIAS [9]; SUS [5]; UMUX [8]; ResQue [12]; BUS [4]. На основе этих субконструктов нами были составлены пункты (см. приложение 1) и выдвинуты следующие гипотезы.

- 1) Конвергентная валидность: значимая положительная корреляция ( $\geq 0,3$ ) с опросником “Оценка удобства использования чат-ботов (BUS-11)” [2]
- 2) Дискриминантная валидность: отсутствие значимой корреляции ( $p > 0,05$ ) с опросником “Шкала удовлетворенности жизнью (SWLS)” [1]
- 3) Содержательная валидность: индекс содержательной валидности ( $CVI \geq 0,75$ )
- 4) Факторная валидность:
  - пункты, относящиеся к одному фактору, будут иметь высокие факторные нагрузки по одному и тому же фактору ( $\geq 0,4$ ), и низкие - по всем остальным
- 1) Надежность: McDonald’s  $\omega$ ,  $\alpha$  Кронбаха  $\geq 0,7$

Пункты были проверены и отредактированы с помощью четырех экспертов в области когнитивной психологии и UX ( $CVI \geq 0,75$ ).

Расчет выборки на основе метода анализа данных корреляции (для конвергентной и дискриминантной валидности) показывает 115 человек (см. приложение 2), однако для качественного факторного анализа, согласно статье Вортингтон и Уиттакер [14], необходимо от 150 человек, а желательно - 200. Поэтому нашим минимальным объемом для выборки является 150 человек.

Тем не менее, на полученных данных был произведен предварительный анализ данных на выборке  $N = 117$  респондентов от 18 до 52 лет ( $M = 21.05$ ,  $SD = 6.18$ ). 71% респондентов – женщины. Для 83 опрошенных последняя полученная ступень образования – среднее общее, для 23 – высшее. Никогда ранее не пользовались Яндекс Лавкой 18% респондентов и 83% никогда ранее не пользовались ИИ-помощником в рамках этого приложения.

Распределения ответов по пунктам всех трех шкал достаточно вариативны, но в основном смещены в сторону высоких значений. Трудность большинства пунктов находится в диапазоне 0.63-0.90 (см. приложение 3), дискриминативность – в диапазоне 0.24-0.70 (см. также приложение 3).

Заметным исключением оказался один пункт из шкалы удобства с трудностью 0.3 и нулевой дискриминативностью: судя по данным, этот пункт (“Я бы не смог(ла) выполнить задачу в этом сервисе без ИИ-помощника”) не связан с конструктом, измеряемым данным инструментом. Дальнейшие показатели вычислялись после исключения данного пункта.

Показатели внутренней согласованности соответствуют стандартным нормам:  $\alpha = 0.84$  ( $\omega = 0.85$ ) для шкалы доверия (см. приложение 4),  $\alpha = 0.79$  ( $\omega = 0.80$ ) для шкалы персонализации (см. приложение 5) и  $\alpha = 0.82$  ( $\omega = 0.83$ ) для шкалы удобства (см. приложение 6). Межпунктовые корреляции являются положительными и находятся в диапазоне от 0.02 до 0.67 (см. приложения 7, 8, 9).

Также было показано, что данные подходят для факторного анализа: критерий сферичности Бартлетта статистически значим ( $[U+1D712]^2(351) = 1604.75$ ,  $p < .001$ ) и индекс КМО равен 0.86, что говорит о хорошей пригодности данных для факторизации. Результаты параллельного анализа подтверждают трехфакторную структуру опросника (см. приложение 10).

В связи с предполагаемой коррелированностью факторов был проведен разведочный анализ с косоугольным вращением *oblimin* (см. приложение 11). Полученная трехфакторная структура только частично соответствует теоретической модели: большая часть пунктов шкалы доверия и все пункты шкалы персонализации распределены между первым и вторым факторами. Часть пунктов шкалы удобства относится к третьему фактору, однако остальные пункты этой шкалы также распределены между первым и вторым факторами, при этом для некоторых пунктов характерны перекрестные нагрузки. Полученные результаты указывают на необходимость дополнительной проработки теоретической

модели и структуры шкал опросника, а также доработки отдельных пунктов, после чего требуется повторная эмпирическая проверка инструмента на расширенной выборке.

Подводя итоги, можно сказать, что опросник для оценки UX ИИ-инструментов был создан, но требует повторного анализа данных на большей выборке и дальнейшего уточнения списка пунктов и структуры в области выделенных факторов, что является целью нашего следующего исследования. //

### Источники и литература

- 1) Осин Е.Н., Леонтьев Д.А. Краткие русскоязычные шкалы диагностики субъективного благополучия: психометрические характеристики и сравнительный анализ // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 1. С. 117–142.
- 2) Рафикова А.С., Воронин А.Н. Оценка удобства использования чат-ботов: адаптация опросника BUS-11 на русскоязычной выборке // Экспериментальная психология. 2025. Т. 18. № 3. С. 194–210.
- 3) Allouch M., Azaria A., Azoulay R. Conversational agents: Goals, technologies, vision and challenges // Sensors. 2021. Vol. 21. No. 24. С. 8448.
- 4) Borsci S., Malizia A., Schmettow M., Van Der Velde F., Tariverdiyeva G., Balaji D., Chamberlain A. The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents // Personal and Ubiquitous Computing. 2022. Vol. 26. No. 1. С. 95-119
- 5) Brooke J. SUS-A quick and dirty usability scale // Usability Evaluation in Industry. 1996. С. 4-7.
- 6) Budhwar P., Malik A., De Silva M.T.T., Thevisuthan P. Artificial intelligence – challenges and opportunities for international HRM: a review and research agenda // The International Journal of Human Resource Management. 2022. Vol. 33. No. 6. С. 1065-1097.
- 7) Fan H., Poole M. Perspectives on personalization // Proceedings of the 9th Americas Conference on Information Systems (AMCIS 2003). Tampa, FL, USA, August 4-6, 2003.
- 8) Finstad K. The usability metric for user experience // Interacting with Computers. 2010. Vol. 22. No. 5. С. 323-327.
- 9) Lee J.D., See K.A. Trust in automation: Designing for appropriate reliance // Human Factors. 2004. Vol. 46. No. 1. С. 50-80.
- 10) Nawaz N., Arunachalam H., Pathi B.K., Gajenderan V. The adoption of artificial intelligence in human resources management practices // International Journal of Information Management Data Insights. 2024. Vol. 4. No. 1. С. 100208.
- 11) Poser M., Küstermann G.C., Tavanapour N., Bittner E.A. Design and evaluation of a conversational agent for facilitating idea generation in organizational innovation processes // Information Systems Frontiers. 2022. Vol. 24. No. 3. С. 771-796.
- 12) Pu P., Chen L., Hu R. A user-centric evaluation framework for recommender systems // Proceedings of the Fifth ACM Conference on Recommender Systems. October 2011. С. 157-164.
- 13) Radziwill N.M., Benton M.C. Evaluating quality of chatbots and intelligent conversational agents. arXiv:1704.04579. 2017.

- 14) Worthington R.L., Whittaker T.A. Scale Development Research: A Content Analysis and Recommendations for Best Practices // The Counseling Psychologist. 2006. Vol. 34. No. 6. С. 806-838.
- 15) ГОСТ Р ИСО 9241-210-2016. Эргономика взаимодействия человек-система. Часть 210. Человеко-ориентированное проектирование интерактивных систем: <https://prot.ect.gost.ru/document.aspx?control=7&id=205428>

### Иллюстрации

Шкала	Номер вопроса	Код вопроса	Тип	Формулировка
1	1	et1	п	Мне понравился стиль общения ИИ-помощника
1	6	et2	о	Мне было неприятно общаться с ИИ-помощником
1	11	et3	п	Ответы ИИ-помощника были доброжелательными
1	17	et4	о	Взаимодействие с ИИ-помощником вызвало отрицательные эмоции
1	22	et5	п	Эмоции от общения с ИИ-помощником были положительными
1	25	et6	о	Ответы ИИ-помощника меня раздражали
1	2	ft1	о	Я настороженно относился/относилась к ответам ИИ-помощника
1	7	ft2	п	Я уверен(а), что ИИ-помощник действовал прозрачно и в интересах пользователя
1	12	ft3	о	Я перепроверял(а) ответы ИИ-помощника
1	18	ft4	п	Я доверял(а) ответам ИИ-помощника
1	24	ft5	п	Я мог(ла) полагаться на ответы ИИ-помощника
2	3	pr1	п	ИИ-помощник хорошо понимал цель моего взаимодействия с ним
2	8	pr2	о	ИИ-помощник часто НЕ включал в ответ важные мне детали
2	13	pr3	п	ИИ-помощник подстраивался в случае изменения моего запроса
2	16	pr4	п	ИИ-помощник давал подходящие мне ответы
2	19	pr5	о	ИИ-помощник плохо учитывал прошлый опыт моего взаимодействия с ним
2	23	pr6	о	ИИ-помощник часто НЕ понимал, чего я хочу
2	27	pr7	п	ИИ-помощник учитывал контекст общения с ним при ответе на запрос
3	4	sp1	о	Мне было тяжело разобраться с тем, как работает ИИ-помощник
3	9	sp2	п	Использовать инструмент было легко и понятно
3	14	sp3	о	Я не понял(а), что умеет делать этот ИИ-помощник
3	20	sp4	п	У меня НЕ возникло трудностей во время использования ИИ-помощника.
3	28	sp5	о	Мне бы пригодилась инструкция по использованию этого ИИ-помощника
3	5	ef1	п	ИИ-помощник ускорил и упростил решение моей задачи.
3	10	ef2	о	Без ИИ-помощника я бы выполнил(а) свою задачу быстрее
3	15	ef3	п	Я бы не смог(ла) выполнить задачу в этом сервисе без ИИ-помощника
3	21	ef4	о	ИИ-помощник меня только запутал
3	26	ef5	п	Благодаря ИИ-помощнику я открыл(а) для себя новые возможности сервиса

Таблица с формулировками пунктов, их порядковыми номерами в опроснике, типами (п – прямой вопрос, о – обратный вопрос) и кодами для анализа данных

Рис. : приложение 1

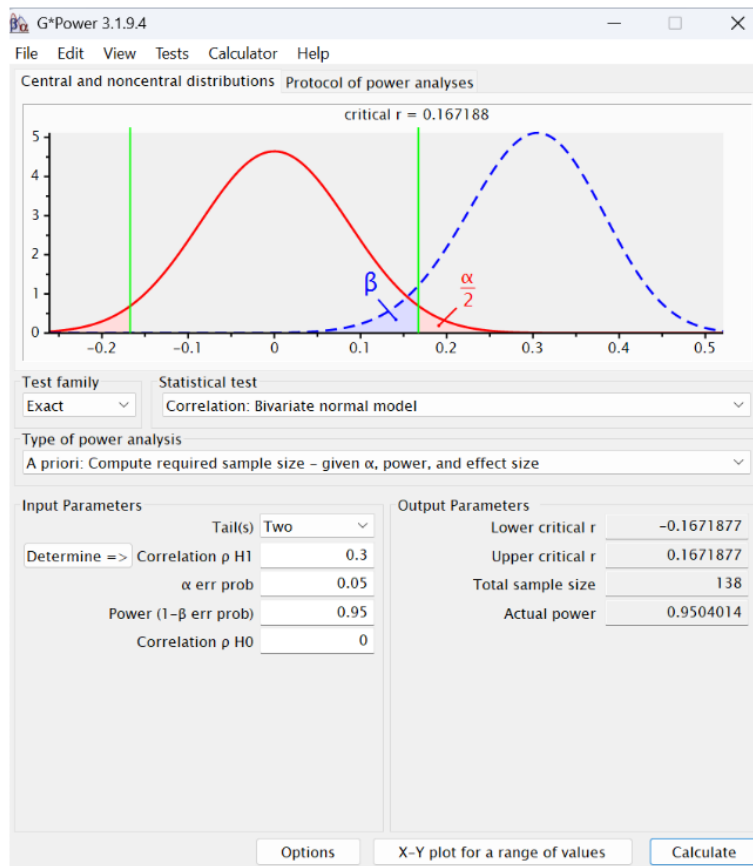


Рис. : приложение 2

Код пункта	Трудность	Дискриминативность
et1	0.759	0.516
et2	0.822	0.579
et3	0.896	0.341
et4	0.843	0.649
et5	0.749	0.694
et6	0.815	0.652
ft1	0.682	0.561
ft2	0.742	0.337
ft3	0.694	0.243
ft4	0.744	0.553
ft5	0.773	0.634
pr1	0.773	0.568
pr2	0.684	0.549
pr3	0.817	0.262
pr4	0.786	0.702
pr5	0.675	0.437
pr6	0.744	0.641
pr7	0.752	0.565
sp1	0.858	0.486
sp2	0.863	0.582
sp3	0.839	0.627
sp4	0.826	0.517
sp5	0.723	0.308
ef1	0.723	0.573
ef2	0.631	0.570
ef3	0.304	-0.091
ef4	0.836	0.627
ef6	0.720	0.494

Таблица значений трудности и дискриминативности для пунктов опросника

Рис. : приложение 3

### Надежность: шкала доверия

Статистики надежности весов

	Кронбах $\alpha$	Макдональдс $\omega$
шкала	0.841	0.850

[3]

Рис. : приложение 4

### Надежность: шкала персонализации

Статистики надежности весов

	Кронбах $\alpha$	Макдональдс $\omega$
шкала	0.785	0.795

[3]

Рис. : приложение 5

### Надежность: шкала удобства

Статистики надежности весов

	Кронбах $\alpha$	Макдональдс $\omega$
шкала	0.819	0.833

[3]

Рис. : приложение 6

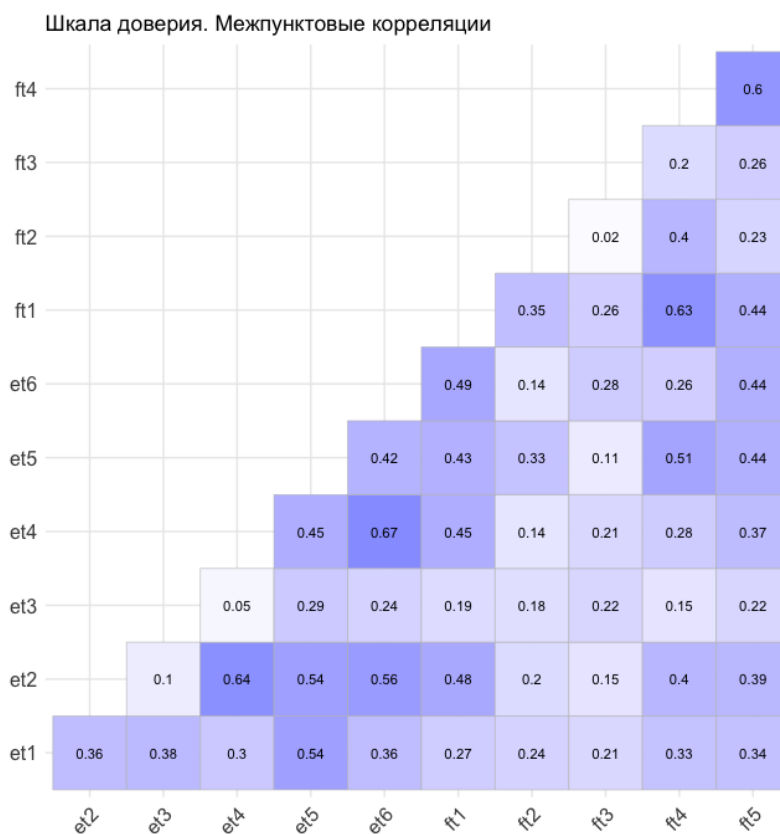


Рис. : приложение 7

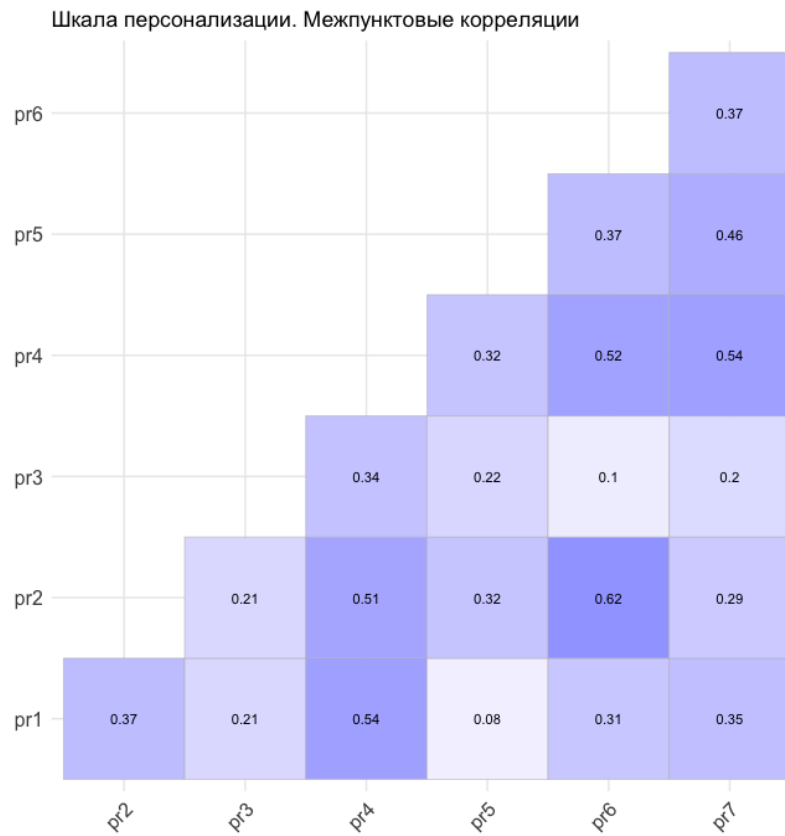


Рис. : приложение 8

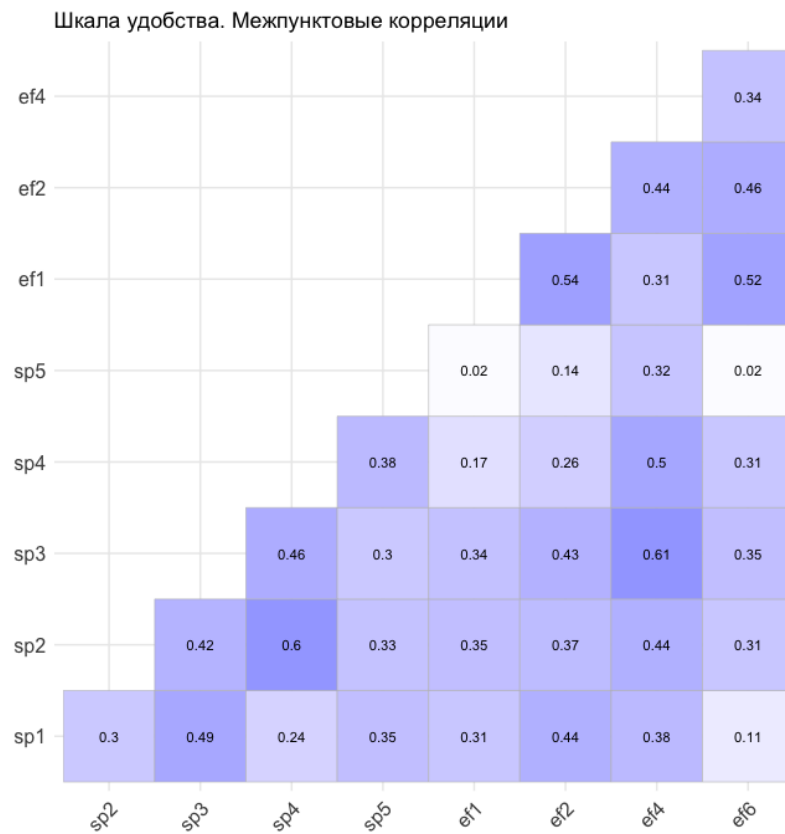


Рис. : приложение 9

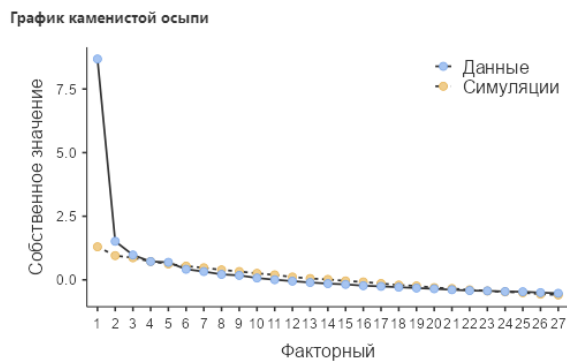


Рис. : приложение 10

### Разведочный факторный анализ

Факторные нагрузки

	Факторный			Уникальность
	1	2	3	
et1	0.547			0.661
et2		0.632		0.508
et3			0.337	0.798
et4		0.770		0.391
et5	0.688			0.417
et6		0.821		0.362
ft1	0.345	0.450	-0.305	0.495
ft2	0.503			0.781
ft3		0.357		0.893
ft4	0.727		-0.318	0.377
ft5	0.538			0.498
pr1	0.718			0.511
pr2		0.387		0.657
pr3				0.851
pr4	0.649			0.413
pr5		0.362		0.783
pr6		0.693		0.448
pr7	0.638			0.574
sp1		0.662		0.560
sp2	0.418		0.394	0.534
sp3		0.482	0.375	0.457
sp4			0.537	0.494
sp5		0.371	0.338	0.739
ef1	0.699			0.518
ef2	0.333			0.633
ef4		0.516		0.489
ef6	0.656			0.591

Примечание. Метод извлечения 'Минимальный остаток' использовался в сочетании с вращением 'облимин'

[3]

Рис. : приложение 11