Responsible Design & Evaluation of a Conversational Agent for a National Careers Service

Marianne Wilson [1], Peter Cruickshank[1], Dimitra Gkatzia[1] and Peter Robertson [1]

[1] Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, Scotland

Abstract

This PhD project applies a research-through-design approach to the development of a conversational agent for a national career service for young people. This includes addressing practical, interactional and ethical aspects of the system. For each aspect, the design process will capture requirements, identify potential solutions and evaluation criteria. Delphi study and Wizard of Oz methods will be used to facilitate participation in the design by both domain experts and young people. Both groups will be involved in the evaluation of a prototype based on the design. The research aims to bring together information retrieval, human-computer interaction and AI ethics research to improve understanding of applied NLP for a domain with complex information navigation and ethical requirements.

Keywords

Responsible development, career support, domain expertise, user-centered design, participatory design, Delphi study, Wizard of Oz

1. Introduction

Effective career decision-making is important for ensuring the prosperity and wellbeing of individuals and national economies. As such, professional services that support young people to develop these abilities are frequently a matter of public policy [1]. Skills Development Scotland (SDS) are responsible for the provision of universal, free to access career services on behalf of the Scottish Government. This research is part of Wilson's ongoing collaborative PhD sponsored by SDS. A research-through-design methodology is being used to develop a framework for the design and evaluation of a conversational agent as part of SDS's services for young people. The aim of this approach is to incorporate both domain expertise and user preferences to maximize the potential usefulness of the system. A user-centered approach will focus on the systems' ability to support young people to complete career related activities or tasks. Career practitioners will determine which career related activities are appropriate for an automated agent. Participatory methods have been prioritized for identifying three categories of requirements: practical, interactional and ethical. A Delphi study with career experts is underway to establish the high-level practical and ethical requirements for the system. These will be refined using a Wizard-of-Oz survey with young people, that will also capture interaction preferences regarding the conversation. Initial findings from the Delphi study identify career-related information navigation as the preferred task focus for the system. Therefore, later phases of the research will include identifying appropriate conversational information retrieval (CIR) methods to meet the design requirements.

The paper provides an overview of existing careers, conversational user experience and AI ethics literature, highlighting implications for this research. It then discusses the research design, current progress and future work required to address both the research and design questions. It is hoped that the methods used will be of interest to researchers working on user-centered applications for information access such as CIR in real-world domains.

2. Context & Research Questions
2.1. Practical Requirements

Career services deliver a wide range of interventions that span career education, information, advice and guidance (CEIAG) [2, 3]. These are often a matter of public policy due to their importance for both individual and national prosperity [1]. SDS policies promote a person-centered approach to CEIAG service design [4]. This reflects the trajectory of career development, which historically focused on approaches for 'matching' individuals to a suitable career [5], that have now been superseded by constructionist and social justice [6] informed paradigms focused on 'life design' [7]. In line with this, SDS's career services aim to support individuals to develop the skills they will need to manage their careers throughout their life [8]. As part of the universal service offer, professional careers advisers lead both classroom-based group activities and one-to-one sessions at key decision points [9]. My World of Work (MyWoW) is an online resource that supports the universal service offer, providing self-service access to a wide range of curated tools and information for young people, carers, and teachers [9]. In addition to the universal services, intensive coaching support is available for young people who have been assessed as in need of this [9]. SDS are under the direct guidance of the Scottish Government, with priorities determined by the Minister for Higher Education, Further Education, Youth Employment and Training [10].

The complexity of the services raises several issues for maximizing the usefulness of the dialogue system. Defining practical requirements will mean identifying an appropriate task focus from the range of information, education and advice interventions that SDS support. This requires consideration of how the conversational agent should be situated relative to existing services. When a specific task has been identified, it will be necessary to ensure that the design is aligned with SDS's approach to career support, and the current services. This includes design dimensions related to: how, when and where young people will access the dialogue system; the complexity and scope of task that it should be designed to support; suitable content and information sources. Domain experts' participation in the research will be essential for achieving this. Given the broad range of information and interventions that career support encompasses, SDS staff have been identified as domain experts. This is due to both their professional training in career development and their personal experience of the design and delivery of SDS' services. As such, the methodology should ensure that domain experts views are incorporated effectively into decisions about system design and evaluation criteria.

These design questions are encapsulated in **RQ1: Which career support tasks could a dialogue system be useful for in the context of SDS's existing services for young people?**

2.2. Interactional Requirements

Given that there are already a variety of routes for accessing career support available to young people, ensuring that the dialogue system provides a positive user experience (UX) is essential for it to be considered a useful contribution to existing services. This means designing systems that do not only meet users' practical requirements, but that 'align' with their affective expectations [11–14]. Conversational agents raise challenges for UX design [15], and, if poorly designed, can lead to a higher cognitive load for users compared to traditional interfaces [16]. The resource intensive nature of human evaluation has led to a focus on scalable, utterance level methods that allow comparison between different systems [17]. However, there has been a recent increase in work that seeks to integrate Human-Computer Interaction and NLP research [18, 19] and an increased focus on UX in design and evaluation research, particularly in Applied NLP across a range of domains [20–23].

However, traditional approaches to UX design, such as prototyping, wireframing [24] and user-modelling [25] are difficult to apply to conversational agents. Therefore, researchers have proposed novel usability scales [26] and design frameworks rooted in conversational analysis principles [27]. Conversational UX aims to optimize at both utterance level and conversation level. Utterance level design decisions focus on aligning the format, structure and word choice to the needs of the user, referred to as *recipient design* [28–30]. Conversation level approaches aim to optimize the number of utterances required to meet the practical and social requirements of the conversations, referred to as

minimization and *repair* [31–33]. CIR research reflects this trend, with the focus design and evaluation of the entire interaction, not single utterances [34–37]. This includes research on the efficacy of specific conversational repair strategies [38, 39]. CIR research that focuses on how to present information in conversational interfaces in alignment with users' needs [40, 41] mirrors recipient design conversational strategies.

Drawing on conversational analysis of human-human interactions, distinct types of conversation have been identified, that each have specific structures and expectations: teaching, ordinary, service, and counselling [27]. The conversation types identified by conversational analysis align with the four aspects of career services: education, information, advice and guidance. This means that RQ1, which aims to scope the practical task for the dialogue system, is a pre-requisite for addressing questions related to conversational UX. Information-seeking conversations between people have also been identified as having distinctive styles and strategies that impact on the quality of the interaction [29]. However, it should be highlighted that the styles of user-computer conversations do not directly mirror human-human conversations [30, 42]. Therefore, it will be necessary for the research to establish users' expectations and preferences regarding the conversational style to design an appropriate conversational user experience. These are summarized in **RQ2: Which methods can be used to deliver a positive conversational user experience in this context?**

2.3. Ethical Requirements

The conversational agent being designed by this research is a form of artificial intelligence (AI), defined as the use of information technology to automate tasks that historically required human intelligence. AI has recently been the subject of a range of public policy interventions [43–47] that have largely focused on how to address potential risks posed by these systems. Many of the risks identified originate from examples of discriminatory outcomes associated with the use of AI in the public sector [48, 49]. These incidents highlight AI systems role in perpetuating a range of inequalities, where the outcomes and mitigation strategies can be mapped to issues including lack of explainability in the system architecture [50], poor management of training data [51, 52] or failure to account for the socio-technical system in which the AI is deployed [53]. Of relevance to this research, is the fact that EU legislation on AI has highlighted education, vocational training and employment as high-risk domains [54]. Similarly, UNICEF have published specific guidance on AI for young people [44], which has been incorporated into the Scottish Government's AI strategy [55]. Within AI, a wide range of harms associated specifically with dialogue systems have been catalogued that range from infringements on individuals' right to equality [56, 57], to global climate impacts [57, 58] .

Given the potential risks of dialogue systems and potential for high impacts within the CEIAG domain, ensuring that the conversational agent design mitigates these concerns effectively is a key requirement for the research. **RQ3: How can the ethical integrity of dialogue system use in this domain be managed effectively?**

3. Methodology

A research-through-design methodology uses the process of artefact design as a means of simultaneously generating knowledge [59]. The design process for the CEIAG conversational agent will follow a standard approach: identify requirements, identify solutions, prototyping and evaluation. The design process will require identifying, evaluating, adapting and creating methods, frameworks and solutions, that are likely to be of value for other applications. Task-focused evaluation predominantly uses standard metrics to compare performance across systems [60]. However, these are inadequate to address current calls for researchers 'to focus on building models that meet people's needs for different tasks, and that can be evaluated on that basis' [61]. While some user-centered evaluation frameworks that include both practical and interactional requirements have been proposed [18, 26], these are focused on service type conversations, with a clearly defined task goal. This research aims to develop and pilot a suitable framework for user centered evaluation in a domain where the task outcomes cannot be clearly defined. This approach incorporates the process of capturing bespoke evaluation criteria in order to both address the research questions, and ensure the usefulness of the system design for its intended use

context within SDS's services for young people. Therefore, domain expertise will be crucial for the definition and evaluation of appropriate practical outcomes needed to address RQ1. Effective collaboration with domain experts is an open issue in data intensive research [62]. This research will include the use of the Delphi method as a means of addressing this problem, as discussed in section 3.2 below. Collaboration with young people to understand specific preferences for the style and content of career-related conversations with automated agents is essential for addressing RQ2. A combined Wizard-of-Oz and survey is planned to elicit interactional requirements, discussed in section 3.3 below.

3.1. Tools for promoting ethical integrity

Some clear ethical requirements can be inferred from the literature review (section 2.3). It is expected that the final results of the Delphi study, and on-going engagement with SDS staff will lead to these continuing to be refined throughout the design process. A range of tools to address RQ3 have been identified from a search of relevant research and policy publications. The EU AI Act takes a risk-based approach, and, therefore, includes guidance on how to self-assess and document potential ethical impacts of an AI system [54]. Model cards are an established, albeit underutilized, tool for addressing issues that originate from mismanagement of training data [52]. Public registers of AI applications [63, 64] exist to support public consultation and monitoring of AI systems. There is a range of guidance on approaches for facilitating consultation and participation in the design and deployment of AI [44, 65, 66]. No single mitigation strategy has yet been tested and validated as reliable for mitigation of possible risks. However, each of these components operate as complement to each other. For example, risk assessments and model cards can be included in public AI registers, AI registers support public awareness and consultation, which in turn supports the identification and monitoring of risks post-deployment.

In addition to tools specific to system design, SDS are required to monitor the impact of all aspects of their service on equality, diversity and inclusion. This includes conducting and documenting robust equality impact assessments and reviews for all policies and services [50]. Guidance on the ethical standards required are available in the form of established professional standards of conduct and ethics [51, 52].

All of these existing best-practice resources for assuring the ethical integrity of the system will be utilized throughout the design process. In addition to addressing RQ3, this will improve understanding of how both AI and domain-specific ethical guidelines can be leveraged in practice to ensure confidence in the ethical integrity of a conversational agent.

3.2. Delphi Study

Co-designing using participatory methods with domain experts has been highlighted as a method for understanding ethical requirements [67]. However, concerns have been raised about the burden that these methods may place on participants, as well as the fact that the outcomes from using these methods may be difficult to incorporate into standard approaches to research and development [65]. A Delphi study is a survey method [68] that can be applied to participatory research [69]. It has been used across a wide range of disciplines as a way to capture expert opinion [70], although it originated as means for forecasting impacts of technologies [68]. The process involves successive rounds of surveys with domain experts. The results of the preceding round are incorporated into each survey, to allow participants to react to the views expressed by their peers. Anonymity is a requirement of the method, to avoid participants who are prestigious or powerful exerting undue influence on the outcome. The main aim of the method is to facilitate structured communication, where online and asynchronous can support the inclusion of geographically dispersed participants [68]. A common misconception is that the outcome of a traditional Delphi should be statistically validated consensus [68]. In fact, a key benefit of the method is the ability to generate rich qualitative data through the pseudo-dialogue between participants, analogous to focus group methods [71]. However, unlike focus groups, the multiple rounds of a Delphi study provide an opportunity to structure data according to specific research aims while the data is being collected. This means that the existing cohort of domain experts have the opportunity to

reject, refine or validate the analysis as it being conducted. As such, the outcomes of a Delphi study are both co-produced with experts, and customized to the needs of the specific research.

The first phase of this research involves a Delphi study with SDS staff, that aims to identify practical and ethical requirements that the system should meet. Using this method as a component in a research-though-design methodology presents an opportunity to evaluate its efficacy as a method for participatory research with domain experts. So far, two of three rounds have been conducted. 23 participants were recruited from SDS staff with experience in practice with young people, service design & policy. Participant engagement has been high, with the first round generating a rich qualitative dataset. The second round of the survey used direct quotes from the first round as Likert scale statements. 22 responses were received for the second round, with clear support for a conversational agent that focused on supporting young people to navigate career-related information, over education, advice or guidance focused options presented. This raises key challenges based in the information retrieval domain. Career-related information is wide ranging and can be difficult for young people to navigate without support [72]. Participants were clear that the conversational agent should not be designed or perceived by users as a means of making a career decision. As such, designing to meet the practical needs of users will require a focus on how to structure and present information in a way that balances the risk of overwhelming users, while encouraging further exploration beyond the dialogue system.

3.3. Wizard of Oz study

A Wizard of Oz (WOZ) method will then be used to establish the interactional requirements. This is a similarly well-established method, that was developed for user-centered NLP research in the 1980s [73]. Participants are asked to use what they believe to be an automated agent, while the system utterances are actually being controlled by a human researcher [74]. This removes the need to develop functional prototypes before detailed user feedback can be solicited, meaning that a wider variety of interaction styles can be tested [75]. It also addresses the issue of the specific conversational styles adopted when interacting with automated systems [42].

The detailed design for the WOZ is dependent on the outcome of the Delphi. It will be conducted with young people who are within SDS's remit. Recruitment will be conducted through SDS network and will aim for a representative sample of race, gender, geographical location, and learning profiles. Achieving this will support accurate completion of the equality impact assessment for the dialogue system.

Data collection with the WOZ will consist of two phases. The first phase will ask direct questions about their preferences. Using the WOZ interface allows presentation of concrete, interactive examples when soliciting participants' views. This will generate survey style data for analysis. The second phase will involve users interacting with what they believe is a functional prototype of a career information chatbot. This will allow the conversational agent to mimic the adoption of a range of conversation types and strategies to understand how users respond to these. Analysis of these transcripts aims to identify requirements for the conversational aspects of the design. This aims to contribute to the current body of research for human-centered approaches to designing conversational agents with a focus on the quality of both the functionality and interaction [20, 27, 35].

3.4. Future Work

The requirements identified from the data collection described above will be documented and assessed using the ethical tools discussed in Section 3.1. A focused literature review will be conducted to identify potential methods for satisfying these in the prototype. Based on the preliminary findings for the Delphi study, this is expected to have a focus on CIR. The requirements will form a key component of the evaluation criteria for the prototype. Evaluation of the prototype will include domain experts, to determine the suitability of the content, and young people to determine if the interaction meets their expectations.

4. Conclusions

The research-through-design approach will result in a case study of the ethical design and evaluation of a conversational agent. The prototype should meet users practical and interactional needs in a domain where task boundaries and objectives are ambiguous. It should be possible to demonstrate that the design aligns with CEIAG ethics. The design process will produce artefacts in the form of expert and user validated requirements, evaluation criteria and a prototype that has been evaluated by experts and users. These are expected to be useful for the design of conversational agents in other domains. The research also aims to evaluate the Delphi method as a structured, participatory approach to meaningfully and efficiently engage domain experts early in the development cycle. The design process requires giving equal attention to practical, interactional and ethical aspects of the conversational agent, and thereby brings together research from information retrieval, human-computer interaction and AI ethics. In so doing, it aims to improve understanding of how this research can be applied to real-world domains.

5. Acknowledgements

6. References

[1]   Watts AG, Sultana RG, "Career Guidance Policies in 37 Countries: Contrasts and Common Themes", International Journal for Educational and Vocational Guidance 4(2007): 105–122,

[2]   Hooley T, Percy C, Alexander R, Exploring Scotland's Career Ecosystem: Evidence to support the Career Review, Skills Development Scotland, 2021.

[3]   Watts AG, National all-age career guidance services: Evidence and issues, British Journal of Guidance and Counselling 38.1 (2010): 31–44.

[4]   The Scottish Government, Scotland's careers strategy: moving forward, 2020.

[5]   Savickas ML, Savickas S, A History of Career Counselling, in: International Handbook of Career Guidance, Springer International Publishing, 2019.

[6]   Thomsen R, The practice portrait – a method for promoting social justice in practice, Journal of the National Institute for Career Education and Counselling, 36.1 (2016):30–37

[7]   Patton W, Recent Developments in Career Theories: The Influences of Constructivism and Convergence, in: Athanasou JA, Esbroeck R Van, editors, International Handbook of Career Guidance, Springer Netherland, 2008.

[8]   Skills Development Scotland, Delivering Scotland's Career Service A Focus on Career Management Skills, 2020.

[9]   CEDEFOP, Inventory of lifelong guidance systems and practices - UK / Scotland, CareersNet national records, 2020.

[10] Hepburn MSP J, Skills Development Scotland - 2022-23 Letter of Guidance, Scottish Government; 2022, doi: https://www.skillsdevelopmentscotland.co.uk/media/49287/sds-governance-2022-23-letter-of-guidance-march-2022.pdf

[11] Skjuve M, Brandzaeg PB, Measuring user experience in chatbots: An approach to interpersonal communication competence, In: Bodrunova SS, Koltsova O, Følstad A, Halpin H, Kolozaridi P, Yuldashev L, et al., (Eds), International Conference on Internet Science INSCI 2018, Cham: Springer, 2019 p, 113–20, doi: https://link.springer.com/chapter/10.1007/978-3-030-17705-8_10

[12] Rese A, Ganster L, Baier D, Chatbots in retailers' customer communication: How to measure their acceptance? Journal of Retailing and Consumer Services 56 (2020):102176. doi: https://doi.org/10.1016/j.jretconser.2020.102176

[13] Følstad A, Brandtzæg PB, Chatbots and the new world of HCI, Interactions, 24.4 (2017):38–42. URL: https://dl.acm.org/doi/10.1145/3085558

[14] Brandtzaeg PB, Følstad A, Why People Use Chatbots, International Conference on Internet Science INSCI 2017, Cham: Springer;2017: 377–92. doi: https://link.springer.com/chapter/10.1007/978-3-319-70284-1_30

[15] Yang Q, Steinfeld A, Rosé C, Zimmerman J, Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design, In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems CHI '20, New York: Association for Computing Machinery, 2020. doi: http://dx.doi.org/10.1145/3313831.3376301

[16] Nguyen QN, Sidorova A, Torres R, User interactions with chatbot interfaces vs, Menu-based interfaces: An empirical study, Computers in Human Behavior 128 (2022). doi: https://doi.org/10.1016/j.chb.2021.107093

[17] Deriu J, Rodrigo A, Otegi A, Echegoyen G, Rosset S, Agirre E, et al, Survey on evaluation methods for dialogue systems, Artificial Intelligence Review 54.1 (2021). doi: https://doi.org/10.1007/s10462-020-09866-x

[18] Heuer H, Buschek D, Methods for the Design and Evaluation of HCI+NLP Systems, in: Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing, 2021, pp 28–34. doi: https://arxiv.org/abs/2102.13461v1

[19] Gupta K, Joshi M, Chatterjee A, Damani S, Narahari N, Agrawal P, Insights from Building an Open-Ended Conversational Agent, in: 1st Workshop on NLP for Conversational AI, Florence: Association for Computational Linguistics; 2019, p, 106–12.

[20] Haugeland IKF, Følstad A, Taylor C, Alexander C, Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design, International Journal of Human-Computer Studies, 161 (2022):102788. doi: https://doi.org/10.1016/j.ijhcs.2022.102788

[21] Montenegro JLZ, da Costa CA, Janssen LP, Evaluating the use of chatbot during pregnancy: A usability study, Healthcare Analytics, 2 (2022):100072. doi: https://doi.org/10.1016/j.health.2022.100072

[22] Liu Y li, Hu B, Yan W, Lin Z, Can chatbots satisfy me? A mixed-method comparative study of satisfaction with task-oriented chatbots in mainland China and Hong Kong, Computers in Human Behavior, 143 (2023): 107716.

[23] Wiethof C, Poser M, Bittner E, Design and Evaluation of an Employee-Facing Conversational Agent in Online Customer Service, in: Pacific Asia Conference on Information Systems, 2022. doi: https://www.researchgate.net/publication/361328764

[24] Yang Q, Cranshaw J, Amershi S, Iqbal ST, Teevan J, Sketching NLP: A case study of exploring the right things to design with language intelligence, in: Conference on Human Factors in Computing Systems - Proceedings, New York: ACM, 2019. doi: https://doi.org/10.1145/3290605.3300415

[25] Følstad A, Brandtzaeg PB, Users' experiences with chatbots: findings from a questionnaire study, Quality and User Experience 5.1 (2020). doi: https://link.springer.com/article/10.1007/s41233-020-00033-2

[26] Borsci S, Malizia A, Schmettow M, van der Velde F, Tariverdiyeva G, Balaji D, et al, The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents, Personal Ubiquitous Computing 26.1 (2022):95–119. doi: https://link.springer.com/article/10.1007/s00779-021-01582-9

[27] Moore RJ, Szymanski MH, Arar R, Ren GJ, Studies in Conversational UX Design , Moore RJ, Szymanski MH, Arar R, Ren GJ, editors, Cham: Springer International Publishing, 2018. doi: http://link.springer.com/10.1007/978-3-319-95579-7

[28] Spillner L, Wenig N, Talk to Me on My Level - Linguistic Alignment for Chatbots, in: Proceedings Of 23rd ACM International Conference On Mobile Human-Computer Interaction (mobile HCI '21), New York: ACM, 2021. doi: https://doi.org/10.1145/3447526.3472050

[29] Thomas P, Czerwinski M, McDuf D, Craswell N, Mark G, Style and alignment in information-seeking conversation, Proceedings of the 2018 Conference on Human Information Interaction and Retrieval (CHIIR 2018), New York: ACM, 2018, pp. 42–51. doi: https://doi.org/10.1145/3176349.3176388

[30] Avgustis I, Shirokov A, Iivari N, "Please Connect Me to a Specialist": Scrutinising 'Recipient Design' in Interaction with an Artificial Conversational Agent, in: Ardito C, Lanzilotti R,

Malizia A, Petrie H, Piccinno A, Desolda G, et al., editors, Human-Computer Interaction -- Interact 2021, Cham: Springer International Publishing; 2021, pp. 155–76. doi: https://link.springer.com/chapter/10.1007/978-3-030-85610-6_10

[31] Benner D, Elshan E, Schöbel S, Janson A, What do you mean? A Review on Recovery Strategies to Overcome Conversational Breakdowns of Conversational Agents, in: Forty-Second International Conference on Information Systems, Austin, 2021.

[32] Liao QV, Geyer W, Muller M, Khazaen Y, Conversational Interfaces for Information Search, in: Fu WT, van Oostendorp H (eds), Understanding and Improving Information Search, Cham: Springer International Publishing, 2020, pp. 267–87. doi: https://link.springer.com/chapter/10.1007/978-3-030-38825-6_13

[33] Hu Y, Qu Y, Maus A, Mutlu B, Polite or Direct? Conversation Design of a Smart Display for Older Adults Based on Politeness Theory, in: Conference on Human Factors in Computing Systems - Proceedings, New York: ACM, 2022. doi: https://doi.org/10.1145/3491102.3517525

[34] Zamani H, Trippas JR, Dalton J, Radlinski F, Conversational Information Seeking An Introduction to Conversational Search, Recommendation, and Question Answering, Boston: now Books; 2022.

[35] Thomas P, Czerwinksi M, Mcduff D, Craswell N, Theories of Conversation for Conversational IR, ACM Transactions on Information Systems 39.4 (2021).

[36] Radlinski F, Craswell N, A theoretical framework for conversational search, in: Proceedings of the 2017 Conference Human Information Interaction and Retrieval (CHIIR 2017), New Yourk: ACM, 2017, pp, 117–26.

[37] Wilson M, Designing Useful Conversational Interfaces for Information Retrieval in Career Decision-Making Support, in: Advances in Information Retrieval Proceedings of European Conference on Information Retrieval Dublin 2023, 2023, pp. 482–488. doi: https://10.1007/978-3-031-28241-6_56

[38] Wang Z, Ai Q, Simulating and Modeling the Risk of Conversational Search, ACM Transactions on Information Systems (TOIS), 40.4 (2022).

[39] Sun Y, Zhang Y, Conversational recommender system, in: 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018, New York: ACM, 2018, pp, 235–44.

[40] Vakulenko S, Savenkov V, de Rijke M, Conversational Browsing, arXiv preprint, 2020. doi: http://arxiv.org/abs/2012.03704

[41] Kiesel J, Meyer L, Potthast M, Stein B, Meta-Information in Conversational Search, ACM Transactions on Information Systems 39.4 (2021).

[42] Elsweiler D, Frummet A, Harvey M, Comparing Wizard of Oz & Observational Studies for Conversational IR Evaluation, Datenbank-Spektrum 20.1 (2020):37–41.

[43] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2021. URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[44] UNICEF, Policy guidance on AI for children, 2021. URL: https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children

[45] European Commission, White Paper on Artificial Intelligence- A European approach to excellence and trust, 2020. URL: https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

[46] Montasari R, National Artificial Intelligence Strategies: A Comparison of the UK, EU and US Approaches with those Adopted by State Adversaries, Advances in Information Security 101 (2023):139–64. doi: https://10.1007/978-3-031-21920-7_7

[47] Ulnicane I, Knight W, Leach T, Stahl BC, Wanjiku WG, Governance of artificial intelligence: Emerging international trends and policy frames, in: Tinnirello M, editor, The Global Politics of Artificial Intelligence, Boca Raton: CRC Press, 2022, pp. 29–55.

[48] Birhane A, The Impossibility of Automating Ambiguity, Artificial Life 1, (2021):44–61. doi: https://doi.org/10.1162/artl_a_00336

[49] Wilson M, Robertson P, Cruickshank P, Gkatzia D, Opportunities and risks in the use of AI in career development practice, Journal of the National Institute for Career Education and Counselling, 48.1 (2022):48–57. doi: https://doi.org/10.20856/jnicec.4807

[50] Rudin C, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence 1.5 (2019):206–15.

[51] Bender EM, Friedman B, Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science, Transactions of the Association for Computational Linguistics 6 (2018):587–604. doi: https://doi.org/10.1162/tacl_a_00041

[52] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al, Model cards for model reporting, in: Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency (FAT '19), New York: ACM, 2019, pp. 220–9. doi: https://doi.org/10.1145/3287560.3287596

[53] Dignum V, Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way, 1st ed, Cham: Springer International Publishing; 2019.

[54] Luciano Floridi, Matthias Holweg, Mariarosaria Taddeo, Javier Amaya Silva, Jakob Mökander, Yuni Wen, capAI A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act, 2023. doi: http://dx.doi.org/10.2139/ssrn.4064091

[55] Scottish Government, Scotland's Artificial Intelligence Strategy, 2021. URL: www.scotlandaistrategy.com

[56] Blodgett SL, Barocas S, Daumé III H, Wallach H, Language (Technology) is Power: A Critical Survey of "Bias" in NLP, in: 58th Annual Meeting of the Association for Computational Linguistics, 2020. doi: https://www.aclweb.org/anthology/

[57] Bender EM, Gebru T, McMillan-Major A, Shmitchell SSG, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), New York: ACM, 2021, pp. 610–623. doi: https://doi.org/10.1145/3442188.3445922

[58] Hershcovich D, Webersinke N, Kraus M, Bingler JA, Leippold M, Towards Climate Awareness in NLP Research, ArXiv, 2022. URL: http://arxiv.org/abs/2205.05071

[59] Zimmerman J, Forlizzi J, Research through design in HCI, in: Ways of Knowing in HCI , Cham: Springer, 2014, pp. 167–89. doi: https://10.1007/978-1-4939-0378-8_8

[60] Järvelin K, Evaluation, in: Ruthven I, Kelly D (eds), Interactive Information Seeking, Behaviour and Retrieval, Facet, 2018, pp. 113–38.

[61] Gebru T, Mitchell M, We warned Google that people might believe AI was sentient, Now it's happening, The Washington Post, 2022 Jun 17.

[62] Sambasivan N, Veeraraghavan R, The Deskilling of Domain Expertise in AI Development, Conference on Human Factors in Computing Systems, in: Proceedings of the 2022 Conference on Human Factors in Computing Systems, New York: ACM, 2022. doi: https://doi.org/10.1145/3491102.3517578

[63] Cath C, Jansen F, Dutch Comfort: The limits of AI governance through municipal registers, arXiv, 2021. doi: https://arxiv.org/abs/2109.02944v1

[64] Haataja M, Fliert L Van De, Rautio P, Public AI Registers: Realising AI transparency and civic participation in government use of AI , Saidot, 2020, URL: https://uploads-ssl.webflow.com/5c8abedb10ed656ecfb65fd9/5f6f334b49d5444079726a79_AI Registers - White paper 1.0.pdf

[65] Birhane A, Isaac W, Prabhakaran V, Diaz M, Elish MC, Gabriel I, et al, Power to the People? Opportunities and Challenges for Participatory AI, in: EAAMO '22: Equity and Access in Algorithms, Mechanisms, and Optimization, New York: ACM, 2022. doi: https://dl.acm.org/doi/10.1145/3551624.3555290

[66] Pierri P, Goodwin K, Zur-Clark A, Waud A, Murray A, Ensuring Full Participation in the Delivery of Scotland's AI Strategy An international case study and rapid literature review, Scottish AI Alliance, 2022.

[67] Showkat D, Baumer EPS, "It's Like the Value System in the Loop": Domain Experts' Values Expectations for NLP Automation, in: Proceedings of the 2022 ACM Designing Interactive Systems Conference: Digital Wellbeing, (DIS 2022), New York: ACM, 2022, pp. 100–22.

[68] Linstone HA, Turoff M, Delphi: A brief look backward and forward, Technological Forecasting and Social Change 78.9 (2011):1712–1719. doi: http://dx.doi.org/10.1016/j.techfore.2010.09.011

[69] Kezar A, Maxey D, The Delphi technique: an untapped approach of participatory research, in International Journal of Social Research Methodology 19.2 (2016):143–160.

[70] Zartha Sossa JW, Halal W, Hernandez Zarta R, Delphi method: analysis of rounds, stakeholder and statistical indicators, Foresight 21.5 (2019):525–44.

[71] Pickard AJ, Delphi study, in: Research Methods in Information, London: Facet, 2013, pp. 149–56.

[72] Milot-Lapointe F, Savard R, Paquette S, Effect of Labour Market Information (LMI): Comparison Between Independent and Assisted Use of LMI, Canadian Journal of Career Development 17.1 (2018):43–52.

[73] Kelley JF, An empirical methodology for writing User-Friendly Natural Language computer applications, in: CHI '83: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1983.

[74] Dahlbäck N, Jönsson A, Ahrenberg L, Wizard of Oz studies, in: Proceedings of the 1st international conference on Intelligent user interfaces - IUI '93 , New York: ACM, 1993, pp. 193–200. doi: https://doi.org/10.1145/169891.169968

[75] Thies IM, Menon N, Magapu S, Subramony M, O'Neill J, How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians, in: Bernhaupt, R and Dalvi, G and Joshi, A and Balkrishan, DK and ONeill, J and Winckler M (eds), Human Computer Interaction - INTERACT 2017, PT I, Cham: Springer International Publishing, 2017, pp. 441–59. doi: https://link.springer.com/chapter/10.1007/978-3-319-67744-6_28