Reflective Diary

Lecture 1 - 11/10 - Interaction with Machine Learning

This first lecture shed some light on the ebbing relationship between Human Computer Interaction (HCI) and Artificial Intelligence (AI) over the years, and provided a much-needed answer to the question: how are both fields even related?

We distinguish three ages in HCI. In the 1980s, the field was closely aligned to AI: interfaces were designed independently of the main system with task-efficiency in mind and a human operator was expected to read the instruction manual prior to interacting with the system. In the 1990s, users and other stakeholders started being integrated in the design process after the realisation that AI models did not yield more usable machines. Since the 2000s, technology has entered every single aspect of our lives and changed the way in which we do are expected to interact with systems. New design goals have been introduced like User Experience (UX) and aesthetics.

Before 2018, research was kept completely separate from AI, but as AI studies mature this is rapidly changing. "A human-centred strategy will bring AI wider acceptance and higher impact by providing products and services that serve human needs" [1]. But what could human-centered AI look like?

In a world where humanity is facing many social challenges (ethnic wars, famine, pandemics, education crisis, climate crisis, etc.), or as Blackwell [2] calls them "problems of imagination", rather than challenges of natural science, [3] asks whether science is the key to our salvation or the maker of our destruction? Blackwell [2] offers that it is rather the lack of imagination in our engineering efforts that should be held responsible for our problems. He puts forward AI as a champion that could help us make art that re-imagines the world and its challenges. Hall [4] believes that all these crises are interrelated and are the result of a larger crisis: the "outgrowth of man having developed a new dimension – the cultural dimension – most of which is hidden from view" and ignored. Could AI help us re-imagine cultures?

Stories sit at the intersection between culture, art, and increasingly, AI. They live for hundreds or thousands of years by being handed down by those who remember them. Stories are the direct product of cultures and they play a central role in communicating them [5]. Through my keen interest in Japanese culture, I came across Naoko Tosa, a Japanese media artist. Her work explores the art of cross-cultural computing by building interactive story-based systems: an interactive movie system [6, 7], a comedy system [8], and an interactive poem system [9]. Studying the intersection between AI and stories, and investigating areas like interactive story-telling [10] could help understand, re-imagine and transcend cultures. Research is currently exploring a whole spectrum of AI applied to stories: creative text generation [11], machine-in-the-loop creative writing [12] and human-AI collaborative writing [13].

Our interactions with and through computers are ever more pervasive, yet fail to mediate the subtleties of communication and culture, causing unnecessary misunderstandings. With the improvement of natural language AI systems, writing interfaces are going beyond simple grammar-checking and spell-checking, offering content suggestions to spark new ideas. These tools could go as far as bridging cultural gaps. For instance, today emails are now sent across distances and cultures. Achieving human-level general intelligence will necessitate communication skills, skills that rely on us creating a more sophisticated technology than that we have today.

References

[1] Ben Shneiderman. "Human-Centered AI: Ensuring Human Control While Increasing Automation". In: *Proceedings of the 5th Workshop on Human Factors in Hypertext*. HUMAN '22. Barcelona, Spain: Association for Computing Machinery, 2022. ISBN: 9781450394017. DOI: 10.1145/3538882.3542790. URL: https://doi.org/10.1145/3538882.3542790.

[2] Alan Blackwell. "Chapter 12: Re-imagining AI to invent more Moral Codes". In: *Moral Codes*. https://moralcodes.pubpub.org/pub/chapter-12. MIT Press, Sept. 2022.

[3] Malcolm Goh. "Science: The Key to Our Salvation or Destruction?" In: *Pioneer Road: Journal of Undergraduate Research* (2021), p. 30.

[4] Edward T. Hall. *The hidden dimension*. Anchor Books/Doubleday, 1990.

[5] Naoko Tosa. *Cross-Cultural Computing: An artist's journey*. Springer London LTD, 2018.

[6] Ryohei Nakatsu and Naoko Tosa. "Toward the realization of interactive movies - inter communication theater: concept and system". In: *Proceedings of IEEE International Conference on Multimedia Computing and Systems* (1997), pp. 71–77.

[7] Ryohei Nakatsu, Naoko Tosa, and Takeshi Ochi. "Interactive Movie System with Multi-Person Participation and Anytime Interaction Capabilities". In: *Proceedings of the Sixth ACM International Conference on Multimedia*. MULTIMEDIA '98. Bristol, United Kingdom: Association for Computing Machinery, 1998, pp. 129–137. ISBN: 0201309904. DOI: 10.1145/290747.290764. URL: https://doi.org/10.1145/290747.290764.

[8] N. Tosa and R. Nakatsu. "Interactive comedy: laughter as the next intelligence system". In: *Proceedings of 2002 International Symposium on Micromechatronics and Human Science*. 2002, pp. 135–138. DOI: 10.1109/MHS.2002.1058023.

[9] Naoko Tosa and Ryohei Nakatsu. ""Interactive Poem System"". In: *Proceedings of the Sixth ACM International Conference on Multimedia*. MULTIMEDIA '98. Bristol, United Kingdom: Association for Computing Machinery, 1998, pp. 115–118. ISBN: 0201309904. DOI: 10.1145/290747.290762. URL: https://doi.org/10.1145/290747.290762.

[10] Linbo Luo et al. "A review of interactive narrative systems and technologies: A training perspective". In: *SIMULATION* 91 (Feb. 2015). DOI: 10.1177/0037549714566722.

[11] Anne-Gwenn Bosser et al. "Poetic or Humorous Text Generation: Jam Event at PFIA2022". In: *13th Conference and Labs of the Evaluation Forum (CLEF 2022)*. Ed. by Guglielmo Faggioli et al. CEUR Workshop Proceedings Working Notes: JokeR: Automatic Wordplay and Humour Translation. Bologna, Italy: CEUR-WS.org, Sept. 2022, pp. 1719–1726. URL: https://hal.archives-ouvertes.fr/hal-03795272.

[12] Elizabeth Clark et al. "Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories". In: *23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: Association for Computing Machinery, 2018, pp. 329–340. ISBN: 9781450349451. DOI: 10.1145/3172944.3172983. URL: https://doi.org/10.1145/3172944.3172983.

[13] Andy Coenen et al. "Wordcraft: a Human-AI Collaborative Editor for Story Writing". In: (2021). DOI: 10.48550/ARXIV.2107.07430. URL: https://arxiv.org/abs/2107.07430.

Lecture 2 - 18/10 - Mixed initiative

This lecture introduced me to the concept of human agency, the ability to say "I did that" when performing an action and witnessing its consequence [1], as an aspect of design I had never considered. Even more interesting I found was the idea that there could exist a metric for it. Intentional binding is a phenomenon where the mind is tricked into thinking that the time interval between a voluntary action and its external sensory consequence is shorter than what it actually is. Conversely, time between a seemingly non-related or involuntary action and its consequence is stretched by the mind. This time interval can be used as a metric to measure a human's sense of agency when performing a task.

One of the most important factors in how individuals engage with technology is their sense of personal agency [2]. Research into input modality is particularly relevant to the question of agency [3]. On-skin interaction, also called Skinput [4] or skin-computing [5], is a technique that uses body landmarks [6] or wearable bio-acoustic sensors on the user's body, to turn the skin into a finger input surface. As it turns out, on-skin input significantly boosts user feelings of agency compared with traditional button-press [7] and touch-pad [8] inputs. Some work has been conducted to understand the mapping from on-skin input to outcomes in the external environment via displays [9], and gaming [10], giving users greater control over external events triggered with their own skin.

Due to the unique form of epidermal systems (intimate integration with the user's body, low cost), they open up a range of opportunities for applications and real-world deployments [11]. In the future, we can envision epidermal systems to be an integral part of our body. As a collective, they could present us with a whole sensing, computing and interaction ecosystem, providing us with more than an accurate description of the state of our body, but with entirely new means of interacting with technology. This seems a particularly likely avenue for the future given the emerging paradigm shift to Internet of Things (IoT) which comes with a vision of the world where electronic devices and sensors all communicate through the internet to "facilitate our lives" [12]. I see many ethical challenges which generally surround health-related IoT [13], but I would imagine this is where we are heading. However, for HCI, it could mean for new levels of usability and feelings of agency for future applications.

On a different topic. Since my project relates to the wider field of AI-powered writing assistance, I started thinking about user agency in that context. Authorship and plagiarism are very sensitive topics in writing [14] and not just legally speaking. A user wants to feel the “I did that” when they finish writing something. How do you find the right balance between wanting to best assist a writer, and granting them enough agency throughout the writing process and a sense of ownership of the final work? Many studies show that agency operates differently in physical and digital environments [15] and intentional binding research calls for an external sensory consequence and has generally been conducted in controlled lab environments [16]. Can we find a way to use intentional binding as a metric for user agency in writing tasks?

References

[1] Patrick Haggard and Manos Tsakiris. "The Experience of AgencyFeelings, Judgments, and Responsibility". In: *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI* 18 (Aug. 2009), pp. 242–246. DOI: 10.1111/j.1467-8721.2009.01644.x.

[2] Gaurav Kumar Rajput. "A review on agency effect in human computer interaction". In: *Asian Journal of Multidimensional Research* 10.10 (2021), pp. 1450–1456.

[3] Patricia Cornelio et al. "The Sense of Agency in Emerging Technologies for Human-Computer Integration: A Review". In: *Frontiers in Neuroscience* 16 (Aug. 2022). DOI: 10.3389/fnins.2022.949138.

[4] Chris Harrison, Desney Tan, and Dan Morris. "Skinput: Appropriating the Body as an Input Surface". In: Jan. 2010, pp. 453–462. DOI: 10.1145/1753326.1753394.

[5] Jürgen Steimle. "On-skin computing". In: *Communications of the ACM* 65 (Apr. 2022), pp. 38–39. DOI: 10.1145/3511668.

[6] Martin Weigel et al. "Skinmarks: Enabling interactions on body landmarks using conformal skin electronics". In: *proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 3095–3105.

[7] David Coyle et al. "I Did That! Measuring Users' Experience of Agency in Their Own Actions". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 2025–2034. ISBN: 9781450310154. DOI: 10.1145/2207676.2208350. URL: https://doi.org/10.1145/2207676.2208350.

[8] Joanna Bergstrom-Lehtovirta et al. "I Really did That: Sense of Agency with Touchpad, Keyboard, and On-skin Interaction". In: Apr. 2018, pp. 1–8. DOI: 10.1145/3173574.3173952.

[9] Joanna Bergstrom-Lehtovirta, Kasper Hornbæk, and Sebastian Boring. "It's a Wrap: Mapping On-Skin Input to Off-Skin Displays". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–11. ISBN: 9781450356206. DOI: 10.1145/

3173574.3174138. URL: https://doi.org/10.1145/3173574.3174138.

[10] Yang Zhang et al. “SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 1491–1503. ISBN: 9781450333627. DOI: 10.1145/2858036.2858082. URL: https://doi.org/10.1145/2858036.2858082.

[11] Aditya Nittala. “From wearable towards epidermal computing”. PhD thesis. Jan. 2021.

[12] Sachin Kumar, Prayag Tiwari, and Mikhail Zymbler. “Internet of Things is a revolutionary approach for future technology enhancement: a review”. In: *Journal of Big Data* 6 (Dec. 2019). DOI: 10.1186/s40537-019-0268-2.

[13] Brent Mittelstadt. “Ethics of the health-related internet of things: a narrative review”. In: *Ethics and Information Technology* 19.3 (2017), pp. 157–175.

[14] P. Bernt Hugenholtz and João Pedro Quintais. “Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?” In: *IIC - International Review of Intellectual Property and Competition Law* (2021).

[15] John E. McEneaney. “Agency Effects in Human–Computer Interaction”. In: *International Journal of Human–Computer Interaction* 29.12 (2013), pp. 798–813. DOI: 10.1080/10447318.2013.777826. eprint: https://doi.org/10.1080/10447318.2013.777826. URL: https://doi.org/10.1080/10447318.2013.777826.

[16] Carl Michael Galang et al. “Studying sense of agency online: Can intentional binding be observed in uncontrolled online settings?” In: *Consciousness and Cognition* 95 (Oct. 2021), p. 103217. DOI: 10.1016/j.concog.2021.103217.

Lecture 3 - 25/10 - Labelling

Labelling can be viewed from two perspectives which come with very different philosophical assumptions. In the eyes of traditional statistical philosophy, observed data describe some underlying natural law [1] and labels provided by the user are intended to capture that law. The human-centred perspective argues that the machine learning systems depend on humans (and their labelling) rather than on physical laws, and are therefore expected to emulate subjective human judgements. For e.g. large language models trained on texts which have been written by humans offer a kind of mechanised plagiarism [2, 3]. Data is often acquired by humans from humans and almost all supervised learning systems share this dependency. There is a clear tension between the ideals of supervised learning and the irregular, noisy, and subjective reality of human-centric systems.

Recent trends in Machine Learning have raised reasonable concerns: ”the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts” [4]. In response, the move to a human-centred machine learning promises to bring human needs and moral values at the heart of technological pursuits [5]. How does a human-centred approach change the way in which machine learning is done? This still seems very much like an open question, but the increased awareness in both the scientific and the global population is a “herald of coming good”.

Talking about a more human-centered approach to machine learning, we can look at improving the experience of data labelers. One student’s project looks at testing his own multi-labelling prototype which would use batch labelling. Previous attempts at improving labelling investigated the idea of data programming, allowing users to write labeling functions to reveal correlations as a starting training data set [6], or semi-automatic labelling techniques [7, 8]. Batching is a recent AI-assisted UX paradigm for labelling. The method helps data labelers by looking at single labels one at a time and applying them to multiple records, which have been partitioned into coherent groups using AI, at the same time. It turns out that batching significantly improves time and accuracy [9]. I would imagine this should also improve the overall experience of labelers: by offering several records simultaneously, they are presented with several reference points at same time, potentially making the labelling of subtle details easier. We saw an example which makes use of a similar idea in the lecture with SorTable [10], which introduces set-wise comparison labelling.

However, like any AI assistance, it can lead to issues of over-reliance which can be harmful because it degrades human-in-the-loop processes. The human plays an important role in monitoring and providing feedback to ensure "all goes well" [11]. Ashktorab et al. [9] explored interactive machine teaching as mitigation technique for over-reliance to frame the relationship between the AI-driven batching system and the user as a collaborative one, in which the labeler works with the recommended batches to improve them for future users. Machine Teaching is an active area of research that sees the machine learning model training process from a human-centered perspective [12]. However, the study [9] did not find a strong effect for the method, revealing shortcomings that indicate clear directions for future work.

References

[1] Leo Breiman. "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)". In: *Statistical Science* 16.3 (2001), pp. 199–231. DOI: 10.1214/ss/1009213726. URL: https://doi.org/10.1214/ss/1009213726.

[2] Alan Blackwell. "Chapter 10: Codes for creativity and surprise". In: *Moral Codes*. https://moralcodes.pubpub.org/pub/chapter-10. MIT Press, Sept. 2022.

[3] Anne Alexander et al. *Ghosts, Robots, Automatic Writing: an AI Level Study Guide*. Cambridge Digital Humanities: Cambridge, PREA, 2021.

[4] Ziad Obermeyer et al. "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464 (2019), pp. 447–453. DOI: 10.1126/science.aax2342. eprint: https://www.science.org/doi/pdf/10.1126/science.aax2342. URL: https://www.science.org/doi/abs/10.1126/science.aax2342.

[5] Stevie Chancellor. *Towards Practices for Human-Centered Machine Learning*. 2022. DOI: 10.48550/ARXIV.2203.00432. URL: https://arxiv.org/abs/2203.00432.

[6] Alexander Ratner et al. "Snorkel: Rapid Training Data Creation with Weak Supervision". In: *Proc. VLDB Endow.* 11.3 (Nov. 2017), pp. 269–282. ISSN: 2150-8097. DOI: 10.14778/3157794.3157797. URL: https://doi.org/10.14778/3157794.3157797.

[7] Meltem Demirkus, James J. Clark, and Tal Arbel. "Robust Semi-Automatic Head Pose Labeling for Real-World Face Video Sequences". In: *Multimedia Tools Appl.* 70.1 (May 2014), pp. 495–523. ISSN: 1380-7501. DOI: 10.1007/s11042-012-1352-1. URL: https://doi.org/10.1007/s11042-012-1352-1.

[8] Szilard Vajda, Yves Rangoni, and Hubert Cecotti. "Semi-automatic ground truth generation using unsupervised clustering and limited manual labeling: Application to handwritten character recognition". In: *Pattern Recognition Letters* 58 (June 2015). DOI: 10.1016/j.patrec.2015.02.001.

[9] Zahra Ashktorab et al. "AI-Assisted Human Labeling: Batching for Efficiency without Overreliance". In: *Proc. ACM Hum.-Comput. Interact.* 5.CSCW1 (Apr. 2021). DOI: 10.1145/3449163. URL: https://doi.org/10.1145/3449163.

[10] Advait Sarkar et al. "Setwise Comparison: Consistent, Scalable, Continuum Labels for Computer Vision". In: May 2016, pp. 261–271. DOI: 10.1145/2858036.2858199.

[11] Ben Shneiderman. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy". In: *CoRR* abs/2002.04087 (2020). arXiv: 2002.04087. URL: https://arxiv.org/abs/2002.04087.

[12] John J. Dudley and Per Ola Kristensson. "A Review of User Interface Design for Interactive Machine Learning". In: *ACM Trans. Interact. Intell. Syst.* 8.2 (June 2018). ISSN: 2160-6455. DOI: 10.1145/3185517. URL: https://doi.org/10.1145/3185517.

Lecture 4 - 01/11 - Program Synthesis

Program synthesis seeks to automatically copy or infer a program that satisfies a user 's intent, which the user expresses via specification and/or by providing illustrative examples. It is a longstanding goal of artificial intelligence research [1, 2, 3] dating back to the 1940s and 50s [4, 5]. But, despite being considered the "holy grail" of AI [6], achieving good program synthesis remains a challenge today [7].

A fundamental idea of program synthesis is that if a user knows how to perform a task on a computer, that should be sufficient to create a program to perform the task. However, much like natural language, the "languages" used by users to conceptualise their intent in an abstract way are ambiguous [8]: "You do the rest". What is "the rest"? What if it is not always the same? How is the program meant to "do"? It becomes essential that the user and system enter a sort of dialogue to discuss any conclusions made by the system and alter them if needed (generalisation step). But how long is a user prepared to dialogue with a system? This is a perfect instance of the Attention Investment model [9]. Automating does save the user's time and effort, but necessitates time and effort.

Program synthesis has direct applications for various classes of users. While at first I mostly thought of applications for non-expert programmers in areas like data science [10] or education [11, 12]. I found that program synthesis has many application for expert programmers. In particular, Software engineers very frequently find themselves in situations where long-term code quality is traded for short-term gain: "not quite right code which we postpone making it right" [13, 14]. This leads to what is known as a technical debt, which can only be re-payed by refactoring [7]. Program synthesis can be used to automate refactoring based on program semantics [15]. Additionally, program synthesis techniques can help check the correctness of code, verify whether programs terminate or not, and prove program robustness [16, 17, 18, 19]. Control engineers can also use program synthesis to automatically produce correct implementations of digital controllers [20, 21, 22].

Modern code editors and IDEs include an auto-completion capabilities: they predict the next likely token in the program based on the previously typed ones like (for e.g. IntelliSense in Microsoft Visual Studio). Program synthesis has the potential to enhance this by automatically completing whole snippets of code instead of single tokens [6]. Recently, GitHub Lab's released

Copilot, a promising Deep Learning based tool which turns natural language prompts into coding suggestions across dozens of languages. Already, many educators in Computer Science are asking: how will teaching CS look like in post-Copilot era? GitHub have recognized the challenge of integrating GitHub Copilot into the classroom, and are committing to partnering closely with the teaching community on how [23]. It will be interesting to see how Copilot affects future curricula.

Program synthesis is achieved through searching within an enormous space of programs to find one that satisfies a given specification, but the size of the program search space quickly becomes intractable [24]. Machine learning techniques have been successfully applied to overcome this problem [25]. Despite their success, machine learning models are often treated as black boxes: verification and interpretability remain a significant challenge [26]. We find ourselves in a context where the leading machine learning models are becoming increasingly opaque and difficult to interpret, while being used to make critical decisions [27]. Can we leverage synthesis-based techniques to improve the quality of machine learning models [28]?

References

[1] Richard J. Waldinger and Richard C. T. Lee. "PROW: A Step Toward Automatic Program Writing". In: *International Joint Conference on Artificial Intelligence*. 1969.

[2] Zohar Manna and Richard Waldinger. "Toward Automatic Program Synthesis". In: *Commun. ACM* 14 (Mar. 1971), pp. 151–165. DOI: 10.1145/362566.362568.

[3] Zohar Manna and Richard Waldinger. "Knowledge and Reasoning in Program Synthesis." In: *Artif. Intell.* 6 (Jan. 1975), pp. 175–208.

[4] B. Jack Copeland. *Alan Turing's Automatic Computing Engine: The Master Codebreaker's Struggle to build the Modern Computer*. Oxford University Press, Apr. 2005. ISBN: 9780198565932. DOI: 10.1093/acprof:oso/9780198565932.001.0001. URL: https://doi.org/10.1093/acprof:oso/9780198565932.001.0001.

[5] J. W. Backus et al. "The FORTRAN Automatic Coding System". In: *Papers Presented at the February 26-28, 1957, Western Joint Computer Conference: Techniques for Reliability*. IRE-AIEE-ACM '57 (Western). Los Angeles, California: Association for Computing Machinery, 1957, pp. 188–198. ISBN: 9781450378611. DOI: 10.1145/1455567.1455599. URL: https://doi.org/10.1145/1455567.1455599.

[6] Sumit Gulwani, Oleksandr Polozov, and Rishabh Singh. "Program Synthesis". In: *Foundations and Trends® in Programming Languages* 4 (Jan. 2017), pp. 1–119. DOI: 10.1561/2500000010.

[7] Cristina David and Daniel Kroening. "Program synthesis: challenges and opportunities". In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 375 (2017).

[8] Navid Yaghmazadeh et al. "SQLizer: query synthesis from natural language". In: *Proceedings of the ACM on Programming Languages* 1 (Oct. 2017), pp. 1–26. DOI: 10.1145/3133887.

[9] Alan Blackwell. "First Steps in Programming: A Rationale for Attention Investment Models". In: *Proceedings - IEEE 2002 Symposia on Human Centric Computing Languages and Environments, HCC 2002* (Jan. 2003). DOI: 10.1109/HCC.2002.1046334.

[10] Maria I. Gorinova et al. "Transforming spreadsheets with Data Noodles". In: *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2016). DOI: 10.1109/vlhcc.2016.7739694.

[11] Sumit Gulwani, Vijay Anand Korthikanti, and Ashish Tiwari. "Synthesizing geometry constructions". In: *ACM SIGPLAN Notices* 46.6 (2011), pp. 50–61.

[12] Sumit Gulwani. "Example-based learning in computer-aided stem education". In: *Communications of the ACM* 57.8 (2014), pp. 70–80.

[13] Ward Cunningham. "The WyCash portfolio management system". In: *ACM SIGPLAN OOPS Messenger* 4.2 (1992), pp. 29–30.

[14] Philippe Kruchten, Robert L Nord, and Ipek Ozkaya. "Technical debt: From metaphor to theory and practice". In: *Ieee software* 29.6 (2012), pp. 18–21.

[15] Cristina David, Pascal Kesseli, and Daniel Kroening. "Kayak: Safe semantic refactoring to java streams". In: *arXiv preprint arXiv:1712.07388* (2017).

[16] C David, D Kroening, and M Lewis. "Unrestricted termination and non-termination proofs for bit-vector programs". In: ESOP. 2015.

[17] Cristina David, Daniel Kroening, and Matt Lewis. "Using program synthesis for program analysis". In: *Logic for programming, artificial intelligence, and reasoning*. Springer. 2015, pp. 483–498.

[18] Pascal Kesseli. "Danger Invariants". In: (2016).

[19] Rahul Sharma and Alex Aiken. "From invariant checking to invariant inference using randomized search". In: *Formal Methods in System Design* 48.3 (2016), pp. 235–256.

[20] Alexandrina L Dumitrescu. *Chemicals in surgical periodontal therapy*. Springer, 2011.

[21] Alessandro Abate et al. "Sound and automated synthesis of digital stabilizing controllers for continuous plants". In: *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*. 2017, pp. 197–206.

[22] Alessandro Abate et al. "Automated formal synthesis of digital controllers for state-space physical plants". In: *International Conference on Computer Aided Verification*. Springer. 2017, pp. 462–482.

[23] Elise Hollowed and Andrew Rice. *GitHub copilot now available for teachers*. Sept. 2022. URL: https://github.blog/2022-09-08-github-copilot-now-available-for-teachers/.

[24] Kensen Shi et al. *CrossBeam: Learning to Search in Bottom-Up Program Synthesis*. Mar. 2022.

[25] Rudy Bunel et al. *Learning to superoptimize programs*. 2016. DOI: 10.48550/ARXIV.1611.01787. URL: https://arxiv.org/abs/1611.01787.

[26] Christoph Molnar, Giuseppe Casalicchio, and Bernd Bischl. *Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges*. Oct. 2020.

[27] Gregory Plumb, Denali Molitor, and Ameet Talwalkar. "Supervised Local Modeling for Interpretability". In: *CoRR* abs/1807.02910 (2018). arXiv: 1807.02910. URL: http://arxiv.org/abs/1807.02910.

[28] Shikhar Singh. "Exploring synergies between program synthesis and machine learning." In: (2022).

Lecture 5 - 08/11 - Visualisation

This entry will first retrace the history of visualisation much as was done in the lecture, and follow with a discussion of the relationship between visualisation, computing and art in aesthetics. Apologies for the very long history, I just got caught up in it and found it very interesting.

As the saying goes, “a picture is worth a thousand words”. Humans have arranged data into tables since the 2nd century C.E. [1], but graphical representations of quantitative data came much later. Prior to the 17th century, data visualisation primarily existed in the realm of cartography. The demand for better visualisations grew with the rise of print, and in 1765, Joseph Priestley published the first timeline charts, which “proved a commercial success and popular sensation” [2]. He was shortly followed by the pioneer of graphical methods of statistics, William Playfair, who inventing the line, area, bar and pie charts, as well as the cycle graph. Where visualisations had previously only been representations of time or space, they introduced new graphics of complex quantified concepts in space.

The study of statistical graphics entered a “Golden Age” [3] as graphics found new uses: understanding social issues (De Fourcroy, 1782; Booth, 1889), tracking disease outbreaks (Snow, 1854), and recording war casualties (Nightingale, 1858; Minard, 1869). In a context of industrial revolution, the government started opening statistical offices and visualisations progressed to many domains. Recognition in the public’s eye grew. The field finally hit a wall in the early 20th century. Friendly et al. [4] describes it as the “modern dark ages” for data visualization. Statisticians had grown increasingly weary of graphics judging them inaccurate over exact numbers in quantification and formal models. Graphic innovation was also awaiting new ideas and new technologies to support the next wave of data visualisation [5].

The “rebirth of data visualisation” [4] was brought on by the emergence of computer processing in the second half of the 20th century. New technologies allowed statisticians to collect, store, and visualise increasingly larger volumes data. In 1962, John W. Turkey published a landmark paper which called for the recognition of data analysis as a legitimate branch of statistics [6], and spent much of his life making data visualisation respectable again. In France, Jacques Bertin laid the foundations of information visualisation in statistics and cartography (Bertin, 1967), introducing visual perception as something that is governed by rules. In time, visualisations have helped convey complex abstractions in an intuitive way by translating them into physical attributes of vision (length, position, size, shape, colour, etc.).

What does this have to do with HCI? Humans interact with computers via visual interfaces, and HCI research seeks to study these interactions and improve their designs. Interfaces are really interactive visualisations, and so they must follow similar design principles as those derived from human perception. Historically, HCI has involved people from many different backgrounds [7, 8, 9, 10]. Their many perspectives have helped generate new ideas [11], but have hindered collaboration across disciplines [12, 13]. To overcome the challenges of collaborative design, Li et al. [14] looked at using visualisation methods.

How can you design computer displays that are as meaningful as possible to human viewers? "Designing an object to be simple and clear takes at least twice as long as the usual way " [15]. Computer displays are such powerful visual appliances that designers need devote extensive effort to balance the demands of many tasks, diverse users, and challenging requirements [16]. As mentioned in the first lecture, recent years have introduced new design goals like aesthetics which place even greater expectation on designers. Aesthetic is the common interest of visualisation, technology and art [17], and, as it turns out, aesthetic considerations in computing positively affect usability and satisfaction [18, 19, 20, 21]. However, much like art and beauty, what users perceive as pleasant or good design or visualisation is highly subjective. A recent study by Leiva et al. [22] made a first attempt at modeling how different user groups perceive web-page aesthetics using a Convolutional Neural Network. This is an interesting future direction for visualisations.

References

[1] Stephen Few. *Data Visualization for Human Perception*. URL: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception.

[2] James R. Beniger. "Media Content as Social Indicators: The Greenfield Index of Agenda-Setting". In: *Communication Research* 5.4 (1978), pp. 437–453. DOI: 10.1177/009365027800500404. eprint: https://doi.org/10.1177/009365027800500404. URL: https://doi.org/10.1177/009365027800500404.

[3] Michael Friendly. "The Golden Age of Statistical Graphics". In: *Statistical Science* 23 (June 2009). DOI: 10.1214/08-STS268.

[4] Michael Friendly and D. Denis. "The roots and branches of statistical graphics". In: *Journal de la Société Française de Statistique* 141 (Jan. 2000).

[5] M. Friendly and D. J. Denis. *Timeline*. 2001. URL: http://www.datavis.ca/milestones/.

[6] Peter McCullagh. "John Wilder Tukey 16 June 1915–26 July 2000". In: *Biographical Memoirs of Fellows of the Royal Society* 49 (2003), pp. 537–555. ISSN: 00804606. URL: http://www.jstor.org/stable/3650242 (visited on 11/30/2022).

[7] John Carroll. "HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science". In: Apr. 2003, pp. -576. ISBN: 1558608087.

[8] Ahmed Seffah, Jan Gulliksen, and Michel C. Desmarais. "An introduction to human-centered software engineering : Integrating usability in the development process". In: 2005.

[9] Christine E. Wania, Michael E. Atwood, and Katherine W. McCain. "Mapping the field of human-computer interaction (HCI)". In: *Proceedings of the American Society for Information Science and Technology* 43.1 (2006), pp. 1–7. DOI: https://doi.org/10.1002/meet.14504301233. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.14504301233. URL: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/meet.14504301233.

[10] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[11] Yvonne Rogers and Helen Sharp. *Interaction Design: Beyond Human-Computer Interaction*. Ed. by Jenny Preece, Yvonne Rogers, and Helen Sharp. 3rd ed. John Wiley, Jan. 2011.

[12] Mieke Haesen et al. "Draw Me a Storyboard: Incorporating Principles and Techniques of Comics to Ease Communication and Artefact Creation in User-Centred Design". In: 2010.

[13] Joanne Mendel. "A taxonomy of models used in the design process". In: *interactions* 19.1 (2012), pp. 81–85.

[14] Jiwei Li, Will Monroe, and Dan Jurafsky. "Understanding Neural Networks through Representation Erasure". In: *CoRR* abs/1612.08220 (2016). arXiv: `1612.08220`. URL: `http://arxiv.org/abs/1612.08220`.

[15] Nelson and H Theodor. "The Home Computer Revolution". In: 1977.

[16] Blackwell Alan. *Visual Representation*. URL: `https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception`.

[17] Salah Ahmed, Abdullah Mahmud, and Kristin Bergaust. "Aesthetics in Human-Computer Interaction: Views and Reviews". In: July 2009, pp. 559–568. ISBN: 978-3-642-02573-0. DOI: `10.1007/978-3-642-02574-7_63`.

[18] Noam Tractinsky, Arielle Katz, and D. Ikar. "What is beautiful is usable". In: *Interact. Comput.* 13 (2000), pp. 127–145.

[19] Donald Norman. "Emotional Design: Why We Love (or Hate) Everyday Things". In: vol. 27. Jan. 2004. ISBN: 0465051359.

[20] Ranjan Maity et al. "A Non-Linear Regression Model to Predict Aesthetic Ratings of On-Screen Images". In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. OzCHI '15. Parkville, VIC, Australia: Association for Computing Machinery, 2015, pp. 44–52. ISBN: 9781450336734. DOI: `10.1145/2838739.2838743`. URL: `https://doi.org/10.1145/2838739.2838743`.

[21] Ranjan Maity, Akshay Madrosiya, and Samit Bhattacharya. "A Computational Model to Predict Aesthetic Quality of Text Elements of GUI". In: *IHCI*. 2015.

[22] Luis Leiva, Morteza Shiripour, and Antti Oulasvirta. "Modeling how different user groups perceive webpage aesthetics". In: *Universal Access in the Information Society* (Aug. 2022). DOI: 10.1007/s10209-022-00910-x.

Lecture 6 - 15/11 - Fairness and bias

The baseline of ethics in AI is legality: the systems you build must obey the laws. Yet many ML research projects and systems transgress them, even in the UK. They use data about people who never gave their consent (for e.g. data scraping), or re-use data in ways other than what was previously agreed. Further, individuals have a legal right to explanation of why a decision was made. If a decision was made as a result of a machine learning system, then an explanation can be requested by the individual. This is especially challenging because current AI models are opaque and very difficult to explain [1]. As a result, it is quite possible that all commercial uses of neural network-based models that make decisions about the lives of people are illegal in the UK.

What can we do differently from a human-centred perspective? Fairness is not enough: computationally fair systems can still be discriminatory (for e.g. Amazon AI recruiting tool which discriminated against women). Machine learning is a way to encode historical practices into predictions about the future. This is literally prejudice, the opposite of progress. No information gets added to the system by AI and ML systems trained on data about society will reflect society's biases and prejudices [2]. Ethical designers must consider the balance of power inherent to their privileged position. Ethical researchers must be collaborators, not saviours: "nothing about me without me" [3].

I am particularly interested in the applications of this to NLP. Large language models can be "prompted" to perform a range of natural language processing tasks, given some examples of the task as input (for e.g. GPT-3 [4]). Often however, these models express unintended kinds of bias, including stereotypical associations [5, 6, 7, 8, 9, 10, 11], or negative sentiment towards specific groups [12]. This is because the language modeling objective used for many applications is for instance "predict the next token on a webpage from the internet" which is different from "follow the user's instructions helpfully and safely" [13, 14]. The language model objective is misaligned. How do we address this bias?

Averting these unintended behaviors is especially important since some of these language models are deployed and used in hundreds of applications. To address the bias in the datasets, we can use methods like informative down-sampling (or up-sampling) of data: for e.g. removing (or adding) sentences to the dataset until gendered pronouns, nouns, adjectives occur equally, to help reduce gender bias. Unfortunately, these approaches not only impair the general performance of the system significantly, with lower BLEU scores, but also often achieve worse performance when assessed using gender-related metrics [15]. Domain adaptation is a further method which consists in building a biased model and then adapt it in a way that reduces bias which was found to give strong and consistent improvements in gender debiasing for e.g. [16]. However, these methods do not claim to fix the bias problem, and various papers point out that addressing data bias is not a "one-and-done exercise" [11, 17] but requires continual monitoring throughout a dataset's life-cycle [18]. Developing better frameworks for debiasing is very much still an open research question that is receiving a lot of attention today [19, 20, 15, 13]. The answers we could find could has serious implications for system design and building in the future.

References

[1] Gregory Plumb, Denali Molitor, and Ameet Talwalkar. "Supervised Local Modeling for Interpretability". In: *CoRR* abs/1807.02910 (2018). arXiv: 1807.02910. URL: http://arxiv.org/abs/1807.02910.

[2] Shalini Kantayya. "Code Bias [Film]". In: (2020).

[3] Valerie Billingham. "Through the patient's eyes". In: *Salzburg seminar session*. Vol. 356. 1998.

[4] Tom B. Brown et al. "Language Models are Few-Shot Learners". In: *CoRR* abs/2005.14165 (2020). arXiv: 2005.14165. URL: https://arxiv.org/abs/2005.14165.

[5] Christine Basta, Marta Ruiz Costa-jussà, and Noe Casas. "Evaluating the Underlying Gender Bias in Contextualized Word Embeddings". In: *CoRR* abs/1904.08783 (2019). arXiv: 1904.08783. URL: http://arxiv.org/abs/1904.08783.

[6] Iz Beltagy, Arman Cohan, and Kyle Lo. "SciBERT: Pretrained Contextualized Embeddings for Scientific Text". In: *CoRR* abs/1903.10676 (2019). arXiv: 1903.10676. URL: http://arxiv.org/abs/1903.10676.

[7] Keita Kurita et al. "Measuring Bias in Contextualized Word Representations". In: *CoRR* abs/1906.07337 (2019). arXiv: 1906.07337. URL: http://arxiv.org/abs/1906.07337.

[8] Emily Sheng et al. "The Woman Worked as a Babysitter: On Biases in Language Generation". In: *CoRR* abs/1909.01326 (2019). arXiv: 1909.01326. URL: http://arxiv.org/abs/1909.01326.

[9] Jieyu Zhao et al. "Gender Bias in Contextualized Word Embeddings". In: *CoRR* abs/1904.03310 (2019). arXiv: 1904.03310. URL: http://arxiv.org/abs/1904.03310.

[10] Haoran Zhang et al. "Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings". In: *CoRR* abs/2003.11515 (2020). arXiv: 2003.11515. URL: https://arxiv.org/abs/2003.11515.

[11] Emily M. Bender et al. "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. Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 610–623. ISBN: 9781450383097. DOI: 10.1145/3442188.3445922. URL: https://doi.org/10.1145/3442188.3445922.

[12] Ben Hutchinson et al. "Social Biases in NLP Models as Barriers for Persons with Disabilities". In: *CoRR* abs/2005.00813 (2020). arXiv: 2005.00813. URL: https://arxiv.org/abs/2005.00813.

[13] Long Ouyang et al. *Training language models to follow instructions with human feedback*. 2022. DOI: 10.48550/ARXIV.2203.02155. URL: https://arxiv.org/abs/2203.02155.

[14] Romal Thoppilan et al. "LaMDA: Language Models for Dialog Applications". In: *CoRR* abs/2201.08239 (2022). arXiv: 2201.08239. URL: https://arxiv.org/abs/2201.08239.

[15] Marcus Tomalin et al. "The practical ethics of bias reduction in machine translation: why domain adaptation is better than data debiasing". In: *Ethics and Information Technology* 23 (2021), pp. 419–433.

[16] Danielle Saunders and Bill Byrne. "Reducing gender bias in neural machine translation as a domain adaptation problem". In: *arXiv preprint arXiv:2004.04498* (2020).

[17] Ben Hutchinson et al. "Towards accountability for machine learning datasets: Practices from software engineering and infrastructure". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, pp. 560–575.

[18] Dirk Hovy and Shrimai Prabhumoye. "Five sources of bias in Natural Language Processing". In: *Language and Linguistics Compass* 15.8 (2021). DOI: 10.1111/lnc3.12432.

[19] Emily Bender and Batya Friedman. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science". In: *Transactions of the Association for Computational Linguistics* 6 (Dec. 2018), pp. 587–604. DOI: 10.1162/tacl_a_00041.

[20] Su Lin Blodgett et al. "Language (Technology) is Power: A Critical Survey of "Bias" in NLP". In: *CoRR* abs/2005.14050 (2020). arXiv: 2005.14050. URL: https://arxiv.org/abs/2005.14050.

Lecture 7 - 22/11 - Explainable AI

A lecture by Simone Stumpf. Explainability is the ability of an AI system to explain itself, and interpretability is the ability of a user to build an appropriate mental model [1] that guides interaction with the system. This includes being able to use, understand and troubleshoot a system successfully. Indeed, greater model interpretability or better explanations can help reveal incompleteness in problem formalization [2], reveal spurious correlations [3] and identify confounding factors that could lead to bias or discrimination in a decision (for e.g. if race or gender are determining features [4]). Further, it can supporting error analyses or feature discovery [5].

Nothing is naturally interpretable, and it really depends on the user and their capabilities: different stakeholders (lay users, domain experts, regulatory agencies, researchers, developers, etc.) might each require different explanations. Understanding how our models work is absolutely critical for high-stakes domains: for e.g. medical systems with lives at stake, judicial systems for bail and parole conditions, commercial systems for loan approval, hiring systems, etc.

The importance of interpretable models is increasingly being recognized [6, 2], and since 2016, work in Explainable AI (XAI) has exploded [7] with a common goal of building appropriate (or calibrated) trust in AI systems. In attempting to build trust, it is often question of what style of explanation to provide (and also to who). Some preliminary studies looked at what explanation styles end-users preferred. Lim et al. [8] found that explanations describing why the system behaved in a certain way resulted in better understanding and encouraged trust, versus explanations that described why the system did not behave a certain way. Stumpft et al. [9] compared keyword, rule-based and similarity explanations and found no clear overall, with a big variability between users. Robust evaluations are needed to drive progress further, but it is so far unclear which evaluation approaches are more suitable [10].

On one hand, global approaches aim to provide a global view of how a model works. However, machine learning models are often so complex that interpretable, trustworthy global explanations are difficult to attain [10]. Recently, a variety of approaches have been proposed for generating local explanations which aim to explain a prediction for an individual instance. Ribeiro et al. [11] proposed LIME (Local Interpretable Model-Agnostic Explanations), a technique that gives explanations of the predictions of any classifier by learning an interpretable model locally around the prediction. Further methods include using gradients to visualize neural networks [12, 13, 5], decomposition approaches [14, 15] and measuring the effect of removing individual words (or features) [16, 17]. Nyugen [10] made a first step towards robustly evaluating different local explanation approaches for text classification models.

Explainability is as old as the topic of AI itself rather than being a problem that arose through AI [18]. However, XAI research is having to "keep pace with applied AI research in order to close the research gaps that could hinder operational deployment." [19]. This goes beyond technological development of new methods but also requires a shift in socio-technical perspectives [20]. There are a lot of remaining challenges. We still don't know when users want explanation, how to provide them at the right time when they need it and how to adequately adjust them to different user types.

References

[1] Donald A Norman. *Five Papers on Human-Machine Interaction.* Tech. rep. California University of San Diego La Jolla Center for Human Information Processing, 1982.

[2] Finale Doshi-Velez and Been Kim. "Towards A Rigorous Science of Interpretable Machine Learning". In: *arXiv e-prints*, arXiv:1702.08608 (Feb. 2017), arXiv:1702.08608. arXiv: `1702.08608 [stat.ML]`.

[3] Lifu Tu et al. "An Empirical Study on Robustness to Spurious Correlations using Pre-trained Language Models". In: *CoRR* abs/2007.06778 (2020). arXiv: `2007.06778`. URL: `https://arxiv.org/abs/2007.06778`.

[4] Artem Domnich and Gholamreza Anbarjafari. "Responsible AI: Gender bias assessment in emotion recognition". In: *arXiv preprint arXiv:2103.11436* (2021).

[5] Malika Aubakirova and Mohit Bansal. "Interpreting Neural Networks to Improve Politeness Comprehension". In: *CoRR* abs/1610.02683 (2016). arXiv: `1610.02683`. URL: `http://arxiv.org/abs/1610.02683`.

[6] Alex A. Freitas. "Comprehensible Classification Models: A Position Paper". In: *SIGKDD Explor. Newsl.* 15.1 (Mar. 2014), pp. 1–10. ISSN: 1931-0145. DOI: `10.1145/2594473.2594475`. URL: `https://doi.org/10.1145/2594473.2594475`.

[7] Zachary Chase Lipton. "The Mythos of Model Interpretability". In: *CoRR* abs/1606.03490 (2016). arXiv: `1606.03490`. URL: `http://arxiv.org/abs/1606.03490`.

[8] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. "Why and Why Not Explanations Improve the Intelligibility of Context-Aware Intelligent Systems". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: Association for Computing Machinery, 2009, pp. 2119–2128. ISBN: 9781605582467. DOI: `10.1145/1518701.1519023`. URL: `https://doi.org/10.1145/1518701.1519023`.

[9] Simone Stumpf et al. "Interacting meaningfully with machine learning systems: Three experiments". In: *International journal of human-computer studies* 67.8 (2009), pp. 639–662.

[10] Dong Nguyen. "Comparing Automatic and Human Evaluation of Local Explanations for Text Classification". In: *North American Chapter of the Association for Computational Linguistics*. 2018.

[11] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *CoRR* abs/1602.04938 (2016). arXiv: 1602.04938. URL: http://arxiv.org/abs/1602.04938.

[12] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. 2013. DOI: 10.48550/ARXIV.1312.6034. URL: https://arxiv.org/abs/1312.6034.

[13] Jiwei Li et al. "Visualizing and Understanding Neural Models in NLP". In: *CoRR* abs/1506.01066 (2015). arXiv: 1506.01066. URL: http://arxiv.org/abs/1506.01066.

[14] Leila Arras et al. "Explaining Predictions of Non-Linear Classifiers in NLP". In: *CoRR* abs/1606.07298 (2016). arXiv: 1606.07298. URL: http://arxiv.org/abs/1606.07298.

[15] Yanzhuo Ding et al. "Visualizing and Understanding Neural Machine Translation". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 1150–1159. DOI: 10.18653/v1/P17-1106. URL: https://aclanthology.org/P17-1106.

[16] David Martens and Foster J. Provost. "Explaining Data-Driven Document Classifications". In: *New York University Stern School of Business Research Paper Series* (2014).

[17] Jiwei Li, Will Monroe, and Dan Jurafsky. "Understanding Neural Networks through Representation Erasure". In: *CoRR* abs/1612.08220 (2016). arXiv: 1612.08220. URL: http://arxiv.org/abs/1612.08220.

[18] Andreas Holzinger et al. "Causability and explainability of artificial intelligence in medicine". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9.4 (2019), e1312.

[19] Trevor Kistan, Alessandro Gardi, and Roberto Sabatini. "Machine learning and cognitive ergonomics in air traffic management: Recent developments and considerations for certification". In: *Aerospace* 5.4 (2018), p. 103.

[20] Christian Meske et al. "Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities". In: *Information Systems Management* 39.1 (2022), pp. 53–63. DOI: 10.1080/10580530.2020.1849465. eprint: https://doi.org/10.1080/10580530.2020.1849465. URL: https://doi.org/10.1080/10580530.2020.1849465.

Lecture 8 - 29/11 - Final Summary

Prior this course I had taken one undergraduate course in HCI. That course introduced me to the basic methods of HCI for design requirements gathering (focus groups, interviews, personas, contextual inquiries), getting fast feedback (card sorting, sketches, wire-frames, mock-ups, prototypes, storyboards, design fictions), running studies (lab studies, think aloud, questionnaires, qualitative analysis) and evaluating (cognitive walk-through, usability report, heuristics). Following the Universal Methods of Design [1], we were tasked with re-designing the Moodle-equivalent used by the university for a single course. After some background research, we picked a specific task and persona, and designed a mock-up to meet its needs. Through several iterations, we ended up producing a high-fidelity prototyping in Figma following the Gestalt principles [2].

I was particularly interested in the topic of accessibility within HCI and learning how many innovations were the result of accessibility-driven research. For instance, the predictive text suggestions in our smartphone keyboards were first developed as an assistive tool for motor-impaired users and poor typists [3]. This is the "curb-cut" effect [4]: accessible systems help everyone, not just people with disabilities. With universities increasingly relying on Virtual Learning Environments, bridging the accessibility gap in policy and implementation is a subject of HCI research [5].

Intelligent user interfaces (IUI) are driven by the goal of improving the UX or usability of user interfaces with the help of AI. These include speech-based interfaces, chat-bots, visual recognition of users and objects, recommender systems, and adaptive user interfaces. For lack of space, I will address only two of these.

The focus of my study was intelligent text entry systems. They have been the subject of extensive research and innovation: from gesture keyboards [6, 7, 8, 9, 10, 11], alternative keyboard layouts [12, 13, 14], to effective key-target resizing [15, 16], sensor-based adaptation [17, 18, 19] and predictive text suggestions [3, 20, 21, 22].

Similarly, with the spread of online streaming (music, videos, films, etc.), recommender systems (RS) which provide suggestions that "relate to various decision-making processes" [23, p.1] have become a very popular area of research [24]. Researchers across many disciplines have contributed extensively with diverse research approaches [25]: from information retrieval (IR), data mining, information security and privacy, to business and marketing. However, the recommender-system community is currently facing a reproducibility crisis, which challenges the common understanding of the state-of-the-art for recommendation tasks [26]. Various attempts have been made to address this issue [27, 28], and recently, Anelli and Bellog [29] established a guideline for researchers to compare future recommender performances.

Predictive text suggestions help many people write more efficiently, but by their nature they affect what we write [30, 31]. Buschek et al. [31] found that users value suggestions for inspiration, and getting unstuck in email composition, and not just speed and accuracy. If inspiration is what users are seeking, then text predictions of least surprise might be the exact opposite of what they want. As Advait mentioned in class, it might be interesting to suggest elements of most surprise instead to support ideation and creation. We need to come up with novel design choices to better support creative writing [32]. I am particularly interested in future research to develop technologies that could support high-level ideation in creative tasks, such as creative writing with a machine in the loop [32, 33], or human-AI collaborations for music composition [34]. Such AI applications are fueled by and fuel for human imagination: "AI is a branch of literature because it is a work of imagination" [35].

References

[1] Bella Martin and Bruce M. Hanington. *Universal Methods of Design: 100 Ways To Research Complex Problems, develop innovative ideas, and Design Effective Solutions*. Rockport Publishers, 2012.

[2] Kurt Koffka. *Principles of Gestalt psychology*. Routledge, 2013.

[3] John J. Darragh, Ian H. Witten, and Mark L. James. "The Reactive Keyboard: A Predictive Typing Aid". In: *Computer* 23.11 (Nov. 1990), pp. 41–49. ISSN: 0018-9162. DOI: 10.1109/2.60879. URL: https://doi.org/10.1109/2.60879.

[4] Eric Bergman. "Toward Accessible". In: *Advances in human-computer interaction* 5 (1995), p. 87.

[5] Rehema Baguma and Maria Wolters. “Making Virtual Learning Environments Accessible to People with Disabilities in Universities in Uganda”. In: *Frontiers in Computer Science* 3 (June 2021), p. 638275. DOI: 10.3389/fcomp.2021.638275.

[6] David J. Ward, Alan F. Blackwell, and David J. C. MacKay. “Dasher—a Data Entry Interface Using Continuous Gestures and Language Models”. In: *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*. UIST ’00. San Diego, California, USA: Association for Computing Machinery, 2000, pp. 129–137. ISBN: 1581132123. DOI: 10.1145/354401.354427. URL: https://doi.org/10.1145/354401.354427.

[7] Shumin Zhai and Per Ola Kristensson. “The Word-Gesture Keyboard: Reimagining Keyboard Interaction”. In: *Commun. ACM* 55.9 (Sept. 2012), pp. 91–101. ISSN: 0001-0782. DOI: 10.1145/2330667.2330689. URL: https://doi.org/10.1145/2330667.2330689.

[8] Xiaojun Bi et al. “Bimanual Gesture Keyboard”. In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST ’12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 137–146. ISBN: 9781450315807. DOI: 10.1145/2380116.2380136. URL: https://doi.org/10.1145/2380116.2380136.

[9] Anders Markussen, Mikkel Rønne Jakobsen, and Kasper Hornbæk. “Vulture: A Mid-Air Word-Gesture Keyboard”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 1073–1082. ISBN: 9781450324731. DOI: 10.1145/2556288.2556964. URL: https://doi.org/10.1145/2556288.2556964.

[10] Shyam Reyal, Shumin Zhai, and Per Ola Kristensson. “Performance and User Experience of Touchscreen and Gesture Keyboards in a Lab Setting and in the Wild”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 679–688. ISBN: 9781450331456. DOI: 10.1145/2702123.2702597. URL: https://doi.org/10.1145/2702123.2702597.

[11] Ouais Alsharif et al. "Long-Short Term Memory Neural Network for Keyboard Gesture Recognition". In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015.

[12] Shumin Zhai and Per Ola Kristensson. "Interlaced QWERTY: Accommodating Ease of Visual Search and Input Flexibility in Shape Writing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. Florence, Italy: Association for Computing Machinery, 2008, pp. 593–596. ISBN: 9781605580111. DOI: 10.1145/1357054.1357149. URL: https://doi.org/10.1145/1357054.1357149.

[13] Xiaojun Bi, Barton A. Smith, and Shumin Zhai. "Quasi-Qwerty Soft Keyboard Optimization". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: Association for Computing Machinery, 2010, pp. 283–286. ISBN: 9781605589299. DOI: 10.1145/1753326.1753367. URL: https://doi.org/10.1145/1753326.1753367.

[14] Mark Dunlop and John Levine. "Multidimensional Pareto Optimization of Touchscreen Keyboards for Speed, Familiarity and Improved Spell Checking". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 2669–2678. ISBN: 9781450310154. DOI: 10.1145/2207676.2208659. URL: https://doi.org/10.1145/2207676.2208659.

[15] Asela Gunawardana, Tim Paek, and Christopher Meek. "Usability Guided Key-Target Resizing for Soft Keyboards". In: *Proceedings of the 15th International Conference on Intelligent User Interfaces*. IUI '10. Hong Kong, China: Association for Computing Machinery, 2010, pp. 111–118. ISBN: 9781605585154. DOI: 10.1145/1719970.1719986. URL: https://doi.org/10.1145/1719970.1719986.

[16] Dmitry Rudchenko, Tim Paek, and Eric Badger. "Text Text Revolution: A Game That Improves Text Entry on Mobile Touchscreen Keyboards". In: *Proceedings of the 9th International Conference on Pervasive Computing*. Pervasive'11. San Francisco, USA: Springer-Verlag, 2011, pp. 206–213. ISBN: 9783642217258.

[17] Shiri Azenkot and Shumin Zhai. "Touch Behavior with Different Postures on Soft Smartphone Keyboards". In: *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '12. San Francisco, California, USA: Association for Computing Machinery, 2012, pp. 251–260. ISBN: 9781450311052. DOI: 10.1145/2371574.2371612. URL: https://doi.org/10.1145/2371574.2371612.

[18] Mayank Goel, Leah Findlater, and Jacob Wobbrock. "WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 2687–2696. ISBN: 9781450310154. DOI: 10.1145/2207676.2208662. URL: https://doi.org/10.1145/2207676.2208662.

[19] Mayank Goel et al. "ContextType: Using Hand Posture Information to Improve Mobile Touch Screen Text Entry". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, 2013, pp. 2795–2798. ISBN: 9781450318990. DOI: 10.1145/2470654.2481386. URL: https://doi.org/10.1145/2470654.2481386.

[20] Andrew Fowler et al. "Effects of Language Modeling and Its Personalization on Touchscreen Typing Performance". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 649–658. ISBN: 9781450331456. DOI: 10.1145/2702123.2702503. URL: https://doi.org/10.1145/2702123.2702503.

[21] Anjuli Kannan et al. "Smart Reply: Automated Response Suggestion for Email". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 955–964. ISBN: 9781450342322. DOI: 10.1145/2939672.2939801. URL: https://doi.org/10.1145/2939672.2939801.

[22] Mia Xu Chen et al. "Gmail Smart Compose: Real-Time Assisted Writing". In: *arXiv e-prints*, arXiv:1906.00080 (May 2019), arXiv:1906.00080. arXiv: 1906.00080 [cs.CL].

[23] Francesco Ricci, Lior Rokach, and Bracha Shapira. "Recommender Systems Handbook". In: vol. 1-35. Oct. 2010, pp. 1–35. DOI: 10.1007/978-0-387-85820-3_1.

[24] Pradeep Singh et al. "Recommender Systems: An Overview, Research Trends, and Future Directions". In: *International Journal of Business and Systems Research* 15 (Jan. 2021), pp. 14–52. DOI: 10.1504/IJBSR.2021.10033303.

[25] Dietmar Jannach et al. *Recommender Systems*. Vol. 24. Jan. 2010. ISBN: 978-0-521-49336-9.

[26] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. "Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19. Copenhagen, Denmark: Association for Computing Machinery, 2019, pp. 101–109. ISBN: 9781450362436. DOI: 10.1145/3298689.3347058. URL: https://doi.org/10.1145/3298689.3347058.

[27] Joeran Beel et al. "Towards reproducibility in recommender-systems research". In: *User Modeling and User-Adapted Interaction* 26 (Mar. 2016), pp. 69–101. DOI: 10.1007/s11257-016-9174-x.

[28] Nikolaos Polatidis et al. "Reproducibility of experiments in recommender systems evaluation". In: May 2018.

[29] Vito Walter Anelli et al. "Top-N Recommendation Algorithms: A Quest for the State-of-the-Art". In: *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, July 2022. DOI: 10.1145/3503252.3531292. URL: https://doi.org/10.1145%2F3503252.3531292.

[30] Kenneth C. Arnold, Krzysztof Z. Gajos, and Adam T. Kalai. "On Suggesting Phrases vs. Predicting Words for Mobile Text Composition". In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: Association for Computing Machinery, 2016, pp. 603–608. ISBN: 9781450341899. DOI: 10.1145/2984511.2984584. URL: https://doi.org/10.1145/2984511.2984584.

[31] Daniel Buschek, Martin Zürn, and Malin Eiband. "The Impact of Multiple Parallel Phrase Suggestions on Email Input and Composition Behaviour of Native and Non-Native English Writers". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445372. URL: https://doi.org/10.1145/3411764.3445372.

[32] Elizabeth Clark et al. "Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories". In: *23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: Association for Computing Machinery, 2018, pp. 329–340. ISBN: 9781450349451. DOI: 10.1145/3172944.3172983. URL: https://doi.org/10.1145/3172944.3172983.

[33] Melissa Roemmele and Andrew S Gordon. "Automated assistance for creative writing with an rnn language model". In: *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. 2018, pp. 1–2.

[34] Minhyang (Mia) Suh et al. "AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445219. URL: https://doi.org/10.1145/3411764.3445219.

[35] Alan Blackwell. "Chapter 12: Re-imagining AI to invent more Moral Codes". In: *Moral Codes*. https://moralcodes.pubpub.org/pub/chapter-12. MIT Press, Sept. 2022.