ENVISION HUMAN-AI PERCEPTUAL ALIGNMENT FROM A MULTIMODAL INTERACTION PERSPECTIVE

Shu Zhong & Marianna Obrist
Department of Computer Science
University College University
United Kingdom
{shu.zhong.21, m.obrist}@ucl.ac.uk

ABSTRACT

Aligning AI with human intent has seen progress, yet perceptual alignment—how AI interprets what we see, hear, feel, or smell—remains underexplored. This paper advocates for expanding perceptual alignment efforts across multimodal sensory modalities, such as touch and olfaction, which are critical for how humans perceive and interpret their environment. We envision AI systems enabling natural, multimodal interactions in everyday contexts, such as selecting clothing that aligns with temperature and texture preferences or rich sensory ambiences that evoke specific sights, sounds, and smells. By advancing multimodal representation learning and perceptual alignment, this work aims to inspire the computer science and human-computer interaction (HCI) communities to design inclusive, human-centred AI systems for everyday, multisensory experiences.

1 INTRODUCTION

Artificial intelligence (AI) has become increasingly embedded in daily life. Aligning AI systems with human values, goals, and intentions has emerged as a key research focus (Hendrycks et al., 2020; Ouyang et al., 2022). Such alignment ensures that AI respects ethical principles, promotes safety, and enhances human well-being(Askell et al., 2021; Dafoe et al., 2021). Although significant advances have been made in aligning AI in areas such as robustness, interpretability, and ethicality (Hendrycks et al., 2020; Pan et al., 2023), a fundamental yet underexplored challenge lies in aligning AI with human sensory understanding—what we term *perceptual alignment*.

Perceptual alignment focuses on AI's ability to process and reason about sensory experiences—such as vision, touch, and smell—in ways that align with human perception (Velasco & Obrist, 2020; Sucholutsky et al., 2023; Zhong et al., 2024b). Sensory modalities are fundamental to how humans experience the world (Pink, 2015; Obrist & Velasco), whether through the aroma and taste of coffee, the rigidity of a steel table, or the flash and rumble of lightning. Most research focuses on vision and auditory processing, leaving underexplored perceptual domains such as touch and olfaction.

Unlike humans, who rely on sensory experiences to navigate the world, AI processes digitalised inputs via mathematical algorithms. This mode of computation has naturally shown a preference to modality that is easy to digitalise such as vision (e.g. RGB pixels for image recognition (Deng et al., 2009)), or auditory signals (e.g. audio waveforms for speech recognition (Radford et al., 2023)). However, AI struggles with sensory modalities where input is more complex to measure or explicitly define to digitise to a universal unit, such as physical tactile sensations or the olfactory (scent) stimuli (Keller & Vosshall, 2016; Lynott & Connell, 2013). However, sensory domains such as tactile sensations and olfactory present unique challenges—there are no standardized methods to represent these inputs digitally (Keller & Vosshall, 2016; Lynott & Connell, 2013). For instance, scents and lack a universal classification language akin to colour names or hex codes, making digitalise them and alignment inherently complex.

The rapid advancements in AI, particularly in recent years, have sparked discussions around Artificial General Intelligence (AGI). Large Language Models (LLMs) have showcased impressive capabilities in diverse downstream tasks (Devlin et al., 2018; Brown et al., 2020; Ouyang et al., 2022), and serve as foundation models for agentic AI systems (Bommasani et al., 2021; Shavit et al., 2023).

We envision AI seamlessly integrating into our day-to-day activities, enabling natural interactions. For example, imagine asking your AI assistant for clothing suggestions suitable for a 16-degree day from your closet. Current systems, like ChatGPT, might recommend a wool sweater—potentially too warm and coarse for your preference (Zhong et al., 2024a). Ideally, a perceptually aligned AI could intuitively adjust its recommendations based on tactile sensations, suggesting something cooler, smoother, yet still comfortably warm. Such an AI might also dynamically shape an immersive ambient environment, effortlessly recreating the sights, sounds, and smells with a simple prompt: *"Set up a focused environment like a coffee shop."*

This position paper advocates for advancing AI's representation learning across all sensory modalities, particularly beyond vision and hearing. We highlight examples from recent work, discuss challenges, and propose opportunities for inclusive, everyday multisensory design. Our call to action targets the computer science and human-computer interaction (HCI) communities to explore perceptual alignment as a pathway to make AI more human-centred in a day-to-day scenario.

2 Background and Motivations

Advances in human-AI perceptual alignment are mostly limited to vision or sound (Lee et al., 2023b; Peter et al., 2023). Here we discuss the underexplored modalities of touch and olfaction.

2.1 How we understand and Digitalise Touch

Touch is fundamental to how humans perceive and interact with objects in their surroundings (Klatzky et al., 1987; 1985). It is a complex, multifaceted sensory modality influenced by factors such as force, texture, and temperature. The sense of touch allows us to differentiate tactile sensations—such as roughness, smoothness, pressure, and warmth—and facilitates day-to-day activities like opening a door or using tools (Lederman & Klatzky, 2009; Tiest, 2010). Additionally, tactile perception plays a central role in object recognition and manipulation. In the HCI community, research has delved into using the "textile hand"—a term referring to the feel, texture, and overall tactile qualities of textiles when touched against the skin (Kawabata & Niwa, 1991; Behery, 2005; Obrist et al., 2013)—as a means to convey touch experiences (Petreca et al., 2013; 2015; Olugbade et al., 2023). Previous studies highlight the advantages of using actual textile samples, including precise control over textural properties, ecological validity, and a gentler presentation. This makes them particularly suitable for tactile evaluation (Guest & Spence, 2003). However, translating tactile experiences into language remains a significant challenge due to inherent ambiguities in semantic descriptions (Rosenkranz & Altinsoy, 2020; Atkinson et al., 2016).

Tactile sensing is critical to understand the environment and facilitate human-robot interaction (HRI). Recent advances in tactile technologies have led to sensors that utilize various transduction methods to convert mechanical stimuli into electrical signals, such as piezoresistive (Chen et al., 2019), capacitive (Mittendorfer & Cheng, 2011), and optical (Piacenza et al., 2020) approaches. Vision-based tactile sensors, such as GelSight, have also emerged, leveraging computer vision to capture high-resolution tactile data (Agarwal et al., 2021). Moreover, there is growing interest in integrating tactile sensing into AI systems to build comprehensive world models, enhancing robots' ability to interpret and interact with their environments (Gao et al., 2023). Touch-Vision-Language dataset was developed containing paired tactile sensor and visual images, alongside human-annotated and model-generated tactile-semantic labels (Fu et al., 2024). Despite this progress, the lack of standardised measurement frameworks for touch research limits its interoperability and alignment with human perception.

2.2 How we understand and Digitalise Smell

Olfaction, or the sense of smell, plays a crucial role in perception and behaviour, influencing our interaction with the environment and overall well-being, often in subtle ways that go unnoticed (Wilson & Stevenson, 2006; Beşevli et al., 2024; Carter et al., 2024). Odours descriptors do not have a tangible physical reference as colours. Scientists can measure the wavelength to distinguish colour, as well as the chemical structures of a scent (Ohloff et al., 2022). However, for the general public without professional training in chemistry-related fields, it is impossible for them to know its odour based on the chemical name. This task is challenging even when individuals sniff the odour with their noses. From a human perspective, unlike colours, which benefit from a universally recognized vocabulary of names (e.g. red) and hex codes, scents lack a standardized language for straightforward and precise identification. People tend to identify the source of the smell, pointing to a tangible physical entity so that others can have a baseline reference of the olfactory experience (Drake & Civille, 2003; Zarzo, 2008). Efforts like the DREAM challenge have attempted to build predictive models for scents based on molecular structures and descriptive languages (Keller et al., 2017), and classification systems like the Fragrance Wheel attempt to categorize scents (Edwards, 2011).

Building on the challenges of scent categorization, recent linguistic and AI research has explored the connection between molecular structures and odour perception or multimodal representation (Lee et al., 2023a; Keller et al., 2017; Lisena et al., 2022; ode, 2020; Ravia et al., 2020). Studies like the DREAM Olfaction Prediction Challenge have aimed to understand how humans perceive different molecules as scents (Keller et al., 2017). Similarly, the Odeuropa project (ode, 2020) has significantly contributed to digital heritage by creating a smell-linguistic odour dataset (Lisena et al., 2022) and launching a multimodal data challenge that integrates vision and text to categorize sniff behaviours in digital heritage (Zinnen et al., 2022). Graph Neural Network (GNN) has also been applied to predict olfactory descriptors from molecular structures (Lee et al., 2023a). Additionally, mainstream AI in this field often uses electronic noses (e-nose) with interdigitated electrode structures and molecular imprinted polymer sensors for detecting specific chemicals such as limonene (Hawari et al., 2012). These developments indicate that with sufficient data, AI models have the potential to match or even exceed human capabilities in olfactory perception.

3 Perceptual Alignment in Touch and Smell

3.1 Current studies in Alignment in Touch

Touch has long posed challenges in robotics. Tactile understanding is critical in robotic manipulation, HRI, and assistive technologies. Yet, the field still lacks universally recognised metrics for evaluating tactile alignment with the diverse sensors used in HRI tasks.

In human-centred studies, tactile tasks are often performed without visual stimuli to prevent visual cues from influencing touch perception (Biocca et al., 2001). Recent research evaluated the agreement between humans and six vision-based Multimodal Large Language Models (MLLMs) in describing the tactile properties of textiles (Zhong et al., 2024b). Findings indicated that the models' descriptions often differed from human descriptions in sentiment and word choice, highlighting a misalignment between how tactile experiences are perceived and expressed by AI versus humans. Further studies looked at how well LLMs encode and interpret human touch in relation to real-world objects like textiles (Zhong et al., 2024a). These investigations revealed limited perceptual alignment exists, with significant variability across different samples and strong output biases present in the models.

The study also revealed biases in the training datasets. A keyword analysis of common datasets, such as WikiText-103 (Merity et al., 2016) and BookCorpus (Zhu et al., 2015), indicated that textiles-related keywords were much less frequent than visual descriptors like colour. This imbalance likely contributes to the misaligned tactile predictions observed in the models. Participants mentioned that their tactile experiences with textiles were poorly reflected in the models' outputs, emphasizing the gap between AI's interpretation of touch and everyday human experiences.

Recent work in Visual-Tactile has begun to address these challenges. For instance, the Touch-Vision-Language (TVL) dataset was introduced, pairing tactile sensor images with corresponding visual imagery and including both human-annotated and model-generated tactile-semantic labels (Fu et al., 2024). This dataset enabled the development of a vision-language-aligned tactile encoder and a TVL model, which generates text descriptions of tactile sensations. The proposed model outperformed existing models, achieving a top-1 tactile-language accuracy of 36.7% and a tactile-vision accuracy of 79.5%. These results indicate that vision mainly contributes to the overall accuracy, the tactile-language mapping remains in its early stages, echoing previous findings (Zhong et al., 2024a;b).

These findings highlight the need for improved datasets, a greater focus on tactile descriptors during model training, and further development of AI that better matches human tactile experiences.

3.2 Current studies in Alignment in Smell

Efforts to align AI systems with human olfactory perception have primarily focused on understanding the relationship between molecular structures and odour perception or developing multimodal representations (Lee et al., 2023a; Keller et al., 2017; Lisena et al., 2022; ode, 2020).

Recent research investigated whether transformer pre-trained on general chemical structures, MoLFormer (Ross et al., 2022), can encode representations that align with human olfactory perception (Taleb et al., 2025). The model was evaluated using multiple datasets, including expert odour labels (Luebke, 2019; Leffingwell, 2005), continuous perceptual ratings (Sagar et al., 2023; Keller et al., 2017), and similarity judgments (Snitz et al., 2013; Ravia et al., 2020). MoLFormer can predict expert labels and similarity judgements. However, none of the models showed a high correlation with continuous perceptual ratings. This suggests that traditional odour descriptors, like “grassy”, may not fully capture the complexity of odour relationships. The findings suggest that pre-trained transformers can capture aspects of human olfactory perception from chemical structures alone, highlighting both their potential and the limitations of current descriptive labels.

In a novel research direction that diverges from the traditional focus on molecular structure, recent studies have explored everyday users’ olfactory experiences by analysing narrative scent descriptions rather than word labels (Zhong et al., 2024c). These investigations assess how well LLMs capture scent-related semantics within their high-dimensional embedding spaces. Findings suggest that the models encode certain scent-related features (e.g. fresh), but their accuracy remains limited and shows biases toward certain odours such as lemon and peppermint. Moreover, comparisons between participant-generated narrative scent descriptions and those generated by generative AI (e.g. GPT-4) have revealed significant discrepancies. AI-generated descriptions often diverge from everyday expressions, relying heavily on professional terminology like “undertone”.

Current findings highlight both the potential and limitations of AI in aligning with human olfactory perception through chemical profiles or narrative descriptions. Integrating diverse data types and richer descriptive expressions may enhance the naturalness and accuracy of AI olfaction.

3.3 Challenges and Opportunities in Alignment

Touch and smell face unique yet overlapping challenges in achieving perceptual alignment with AI. Unlike vision or hearing, which benefit from established metrics and abundant datasets, these senses are underrepresented in current research. The primary obstacles include the absence of **standardised digital representations**, **scarcity of data**, and **limited descriptive language**.

Data scarcity remains a major barrier. Tactile datasets, for instance, are far less developed than their visual or linguistic counterparts. Although vibration libraries have been introduced to convey tactile feedback (Seifi et al., 2015), comprehensive tactile databases are still lacking. Similarly, existing olfactory datasets, such as molecular odour databases (Keller et al., 2017; Luebke, 2019; Lisena et al., 2022; Lee et al., 2023a), primarily rely on chemical data that often disconnect from everyday olfactory experiences, failing to fully capture the diversity profile of scents. Furthermore, narrative descriptions of tactile and olfactory sensations are rarely documented, despite these experiences being deeply rooted in shared human understanding (common sense).

Advances in foundation models, e.g. LLMs, offer a promising starting point for exploring perceptual alignment. The emergence of haptic (Fleck et al., 2025) and olfactory devices (Hopper et al., 2024) is beginning to integrate these sensory modalities into the predominantly audiovisual digital ecosystem (Cornelio et al., 2023), opening new avenues for digital representation. Future research could further explore how these sensory modalities are **understood, measured, and represented**.

4 Conclusion

Perceptual alignment in AI remains a critical and underexplored area, particularly for sensory modalities like touch and smell. These senses are fundamental to how humans experience and interpret the world but remain challenging to digitise, standardise, and align with AI systems. This paper calls on the computer science and HCI communities to prioritise advancing multimodal representation learning, fostering perceptual alignment for natural multimodal human-AI interactions.

REFERENCES

Smell Heritage – Sensory Mining - Odeuropa, February 2020. URL https://odeuropa.eu.

Arpit Agarwal, Timothy Man, and Wenzhen Yuan. Simulation of vision-based tactile sensors using physics based rendering. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–7. IEEE, 2021.

Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, et al. A general language assistant as a laboratory for alignment. *arXiv preprint arXiv:2112.00861*, 2021.

Douglas Atkinson, Sharon Baurley, Bruna Beatriz Petreca, Nadia Bianchi-Berthouze, and Penelope Watkins. The tactile triangle: a design research framework demonstrated through tactile comparisons of textile materials. *Journal of Design Research*, 14(2):142–170, 2016.

Hassan Behery. *Effect of mechanical and physical properties on fabric hand*. Elsevier, 2005.

Ceylan Beşevli, Giada Brianza, Christopher Dawes, Nonna Shabanova, Sanjoli Mathur, Matt Lechner, Emanuela Maggioni, Duncan Boak, Carl Philpott, Ava Fatah Gen. Schieck, et al. Smell above all: Envisioning smell-centred future worlds. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pp. 2530–2544, 2024.

Frank Biocca, Jin Kim, and Yung Choi. Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators & Virtual Environments*, 10(3):247–265, 2001.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Anna Carter, Ceylan Beşevli, Christopher Dawes, and Marianna Obrist. Scent and space: Navigating smell interactions in public, private, and virtual spaces. *Interactions*, 31(6):38–43, 2024.

Minrui Chen, Weifeng Luo, Zhongqi Xu, Xueping Zhang, Bo Xie, Guanghou Wang, and Min Han. An ultrahigh resolution pressure sensor based on percolative metal nanoparticle arrays. *Nature communications*, 10(1):4024, 2019.

Patricia Cornelio, Chi Thanh Vi, Giada Brianza, Emanuela Maggioni, and Marianna Obrist. Smell and taste-based interactions enabled through advances in digital technology. In *Handbook of Human Computer Interaction*, pp. 1–31. Springer, 2023.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Cooperative ai: machines must learn to find common ground. *Nature*, 593(7857):33–36, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Mary Ann Drake and Gail Vance Civille. Flavor lexicons. *Comprehensive reviews in food science and food safety*, 2(1):33–40, 2003.

Michael Edwards. *Fragrances of the World: Parfums Du Monde: 2010*. Fragrances of the World, 2011.

Joshua J Fleck, Zane A Zook, Janelle P Clark, Daniel J Preston, Darren J Lipomi, Claudio Pacchierotti, and Marcia K O'Malley. Wearable multi-sensory haptic devices. *Nature Reviews Bioengineering*, pp. 1–15, 2025.

Letian Fu, Gaurav Datta, Huang Huang, William Chung-Ho Panitch, Jaimyn Drake, Joseph Ortiz, Mustafa Mukadam, Mike Lambeta, Roberto Calandra, and Ken Goldberg. A touch, vision, and language dataset for multimodal alignment. In *International Conference on Machine Learning*, pp. 14080–14101. PMLR, 2024.

Ruohan Gao, Yiming Dou, Hao Li, Tanmay Agarwal, Jeannette Bohg, Yunzhu Li, Li Fei-Fei, and Jiajun Wu. The objectfolder benchmark: Multisensory learning with neural and real objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17276–17286, 2023.

Steve Guest and Charles Spence. Tactile dominance in speeded discrimination of textures. *Experimental brain research*, 150:201–207, 2003.

Huzein Fahmi Hawari, Nurul Maisyarah Samsudin, Mohd Noor Ahmad, Ali Yeon Md Shakaff, Supri A Ghani, Yufridin Wahab, and Uda Hashim. Recognition of limonene volatile using interdigitated electrode molecular imprinted polymer sensor. In *2012 Third International Conference on Intelligent Systems Modelling and Simulation*, pp. 723–726. IEEE, 2012.

Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, and Jacob Steinhardt. Aligning ai with shared human values. In *International Conference on Learning Representations*, 2020.

Richard Hopper, Daniel Popa, Emanuela Maggioni, Devarsh Patel, Marianna Obrist, Basile Nicolas Landis, Julien Wen Hsieh, and Florin Udrea. Multi-channel portable odor delivery device for self-administered and rapid smell testing. *Communications Engineering*, 3(1):141, 2024.

S Kawabata and Masako Niwa. Objective measurement of fabric mechanical property and quality: its application to textile and clothing manufacturing. *International Journal of Clothing Science and Technology*, 3(1):7–18, 1991.

Andreas Keller and Leslie B Vosshall. Olfactory perception of chemically diverse molecules. *BMC neuroscience*, 17:1–17, 2016.

Andreas Keller, Richard C Gerkin, Yuanfang Guan, Amit Dhurandhar, Gabor Turu, Bence Szalai, Joel D Mainland, Yusuke Ihara, Chung Wen Yu, Russ Wolfinger, et al. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355(6327):820–826, 2017.

Roberta L Klatzky, Susan J Lederman, and Victoria A Metzger. Identifying objects by touch: An "expert system". *Perception & psychophysics*, 37:299–302, 1985.

Roberta L Klatzky, Susan J Lederman, and Catherine Reed. There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of experimental psychology: general*, 116(4):356, 1987.

S. J. Lederman and R. L. Klatzky. Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459, October 2009. ISSN 1943-393X. doi: 10.3758/APP.71.7.1439.

Brian K Lee, Emily J Mayhew, Benjamin Sanchez-Lengeling, Jennifer N Wei, Wesley W Qian, Kelsie A Little, Matthew Andres, Britney B Nguyen, Theresa Moloy, Jacob Yasonik, et al. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023a.

Jiyoung Lee, Seungho Kim, Seunghyun Won, Joonseok Lee, Marzyeh Ghassemi, James Thorne, Jaeseok Choi, O-Kil Kwon, and Edward Choi. Visalign: Dataset for measuring the alignment between ai and humans in visual perception. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023b.

John C Leffingwell. Leffingwell & associates. *Chirality & Odour Perception. Available online: http://www. leffingwell. com/chirality/chirality. htm (accessed on 15 November 2023)*, 2005.

Pasquale Lisena, Daniel Schwabe, Marieke van Erp, Raphaël Troncy, William Tullett, Inger Leemans, Lizzie Marx, and Sofia Colette Ehrich. Capturing the semantics of smell: the odeuropa data model for olfactory heritage information. In *European Semantic Web Conference*, pp. 387–405. Springer, 2022.

W Luebke. The good scents company information system. *Online Access: http://www. thegoodscentscompany. com*, 2019.

Dermot Lynott and Louise Connell. Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, 45:516–526, 2013.

Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. Pointer sentinel mixture models, 2016.

Philipp Mittendorfer and Gordon Cheng. Humanoid multimodal tactile-sensing modules. *IEEE Transactions on robotics*, 27(3):401–410, 2011.

Marianna Obrist and Carlos Velasco. Multisensory experiences: Formation, realization, and responsibilities. *Communications of the ACM*.

Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. Talking about tactile experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1659–1668, 2013.

Günther Ohloff, Wilhelm Pickenhagen, Philip Kraft, and Fanny Grau. *Scent and chemistry: the molecular world of odors*. John Wiley & Sons, 2022.

Temitayo Olugbade, Lili Lin, Alice Sansoni, Nihara Warawita, Yuanze Gan, Xijia Wei, Bruna Petreca, Giuseppe Boccignone, Douglas Atkinson, Youngjun Cho, et al. Fabrictouch: a multimodal fabric assessment touch gesture dataset to slow down fast fashion. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–8. IEEE, IEEE Computer Society, 2023. ISBN 979-8-3503-2743-4. doi: 10.1109/ACII59096.2023.10388086.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35: 27730–27744, 2022.

Alexander Pan, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside, Hanlin Zhang, Scott Emmons, and Dan Hendrycks. Do the rewards justify the means? measuring trade-offs between rewards and ethical behavior in the machiavelli benchmark. In *International Conference on Machine Learning*, pp. 26837–26867. PMLR, 2023.

Silvan David Peter, Shreyan Chowdhury, Carlos Eduardo Cancino-Chacón, and Gerhard Widmer. Are we describing the same sound? an analysis of word embedding spaces of expressive piano performance. *arXiv preprint arXiv:2401.02979*, 2023.

Bruna Petreca, Nadia Bianchi-Berthouze, Sharon Baurley, Penelope Watkins, and Douglas Atkinson. An embodiment perspective of affective touch behaviour in experiencing digital textiles. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 770–775. IEEE, 2013.

Bruna Petreca, Sharon Baurley, and Nadia Bianchi-Berthouze. How do designers feel textiles? In *2015 International conference on affective computing and intelligent interaction (ACII)*, pp. 982–987. IEEE, 2015.

Pedro Piacenza, Keith Behrman, Benedikt Schifferer, Ioannis Kymissis, and Matei Ciocarlie. A sensorized multicurved robot finger with data-driven touch sensing via overlapping light signals. *IEEE/ASME Transactions on Mechatronics*, 25(5):2416–2427, 2020.

Sarah Pink. Doing sensory ethnography. 2015.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pp. 28492–28518. PMLR, 2023.

Aharon Ravia, Kobi Snitz, Danielle Honigstein, Maya Finkel, Rotem Zirler, Ofer Perl, Lavi Secundo, Christophe Laudamiel, David Harel, and Noam Sobel. A measure of smell enables the creation of olfactory metamers. *Nature*, 588(7836):118–123, 2020.

Robert Rosenkranz and M Ercan Altinsoy. Mapping the sensory-perceptual space of vibration for user-centered intuitive tactile design. *IEEE Transactions on Haptics*, 14(1):95–108, 2020.

Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4(12):1256–1264, 2022.

Vivek Sagar, Laura K Shanahan, Christina M Zelano, Jay A Gottfried, and Thorsten Kahnt. High-precision mapping reveals the structure of odor coding in the human brain. *Nature neuroscience*, 26(9):1595–1602, 2023.

Hasti Seifi, Kailun Zhang, and Karon E MacLean. Vibviz: Organizing, visualizing and navigating vibration libraries. In *2015 IEEE World Haptics Conference (WHC)*, pp. 254–259. IEEE, 2015.

Yonadav Shavit, Sandhini Agarwal, Miles Brundage, Steven Adler, Cullen O'Keefe, Rosie Campbell, Teddy Lee, Pamela Mishkin, Tyna Eloundou, Alan Hickey, et al. Practices for governing agentic ai systems. *Research Paper, OpenAI, December*, 2023.

Kobi Snitz, Adi Yablonka, Tali Weiss, Idan Frumin, Rehan M Khan, and Noam Sobel. Predicting odor perceptual similarity from odor structure. *PLoS computational biology*, 9(9):e1003184, 2013.

Ilia Sucholutsky, Lukas Muttenthaler, Adrian Weller, Andi Peng, Andreea Bobu, Been Kim, Bradley C. Love, Erin Grant, Iris Groen, Jascha Achterberg, Joshua B. Tenenbaum, Katherine M. Collins, Katherine L. Hermann, Kerem Oktar, Klaus Greff, Martin N. Hebart, Nori Jacoby, Qiuyi Zhang, Raja Marjieh, Robert Geirhos, Sherol Chen, Simon Kornblith, Sunayana Rane, Talia Konkle, Thomas P. O'Connell, Thomas Unterthiner, Andrew K. Lampinen, Klaus-Robert Müller, Mariya Toneva, and Thomas L. Griffiths. Getting aligned on representational alignment, 2023.

Farzaneh Taleb, Miguel Vasco, Antonio Ribeiro, Mårten Björkman, and Danica Kragic. Can transformers smell like humans? *Advances in Neural Information Processing Systems*, 37:72032–72060, 2025.

Wouter M Bergmann Tiest. Tactual perception of material properties. *Vision research*, 50(24): 2775–2782, 2010.

Carlos Velasco and Marianna Obrist. *Multisensory experiences: Where the senses meet technology*. Oxford University Press, 2020.

Donald A Wilson and Richard J Stevenson. *Learning to smell: olfactory perception from neurobiology to behavior*. JHU Press, 2006.

Manuel Zarzo. Relevant psychological dimensions in the perceptual space of perfumery odors. *Food Quality and Preference*, 19(3):315–322, 2008. doi: https://doi.org/10.1016/j.foodqual.2007.10.007.

Shu Zhong, Elia Gatti, Youngjun Cho, and Marianna Obrist. Exploring human-ai perception alignment in sensory experiences: Do llms understand textile hand? *arXiv preprint arXiv:2406.06587*, 2024a.

Shu Zhong, Elia Gatti, Youngjun Cho, and Marianna Obrist. Feeling textiles through ai: An exploration into multimodal language models and human perception alignment. In *Proceedings of the 26th International Conference on Multimodal Interaction*, pp. 33–37, 2024b.

Shu Zhong, Zetao Zhou, Christopher Dawes, Giada Brianz, and Marianna Obrist. Sniff ai: Is my'spicy'your'spicy'? exploring llm's perceptual alignment with human smell experiences. *arXiv preprint arXiv:2411.06950*, 2024c.

Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

Mathias Zinnen, Prathmesh Madhu, Ronak Kosti, Peter Bell, Andreas Maier, and Vincent Christlein. Odor: The icpr2022 odeuropa challenge on olfactory object recognition. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pp. 4989–4994. IEEE, 2022.