SPRITE+ Explainer #003

Series Editor: Mark Elliot

AI and Identity Erosion

By Yang Lu, Cigdem Sengul, Edward Apeh, Iain Reid, John McAlaney, Kate Han and Pejman Saeghe

Artificial intelligence (AI) is not only transforming economies and societies but also reshaping how people see themselves. As algorithms begin to write, design, decide, and even empathise, the unique qualities that once anchored human identity risk being overshadowed or devalued. This explainer explores how AI can erode personal, professional, and social identity, and what can be done to preserve the roles, narratives, and self-concepts that make us human.

What is Identity?

The notion of identity has received extensive research attention across the social sciences. In this explainer, we consider human identity encompassing the unique qualities, characteristics, and beliefs that define a person, shaping how they see themselves and how others perceive them, in keeping with social psychological research [2]. Individuals can have multiple self-identities depending on the social context in which their identity is invoked and located. This is as predicted by social psychological and sociological theories, including:

- Symbolic Interactionism and Dramaturgical Theory [14, 3] view identity not only personal but also sociological, Identities are socially constructed and performed through interaction, which can be shaped by infrastructures, technologies, and social expectations [10].
- Social Identity Theory [26] suggests that part of an individual's self-concept derives from membership of particular groups and the social value attached to those groups.
- Role Identity Theory [24] emphasises that identity is shaped by the roles individuals perform in society and the expectations attached to those roles.
- Identity Work and Narrative Theory [13,21] focuses on the ways people maintain identity through the stories they tell about themselves and their lives.
- Networked Self theory [17] states that in digital contexts individuals create multiple and overlapping identities across platforms, including private and public selves.

Finally, **digital Identity** as defined in [8] refers to electronic representations used to authenticate and distinguish individuals in digital interactions. Unlike traditional forms of identity that are grounded in social roles or a composite of attributes (e.g., names, email addresses, usernames, biometric data, or digital certificates) that enable participation in online spaces, digital identities are not static but are constructed, managed, and negotiated across platforms and organisations. They extend and reshape human identity by engendering trust and enabling privacy and access

AI and Identity Erosion

In this explainer, we adopt the UK Data Ethics Framework's definition of AI: 'AI can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence [9]. Advances in AI are reshaping the way humans work, develop, and interact. While AI technologies have become increasingly capable across a wide range of tasks [12], they also risk eroding core elements of human identity.

Identity erosion describes the loss of self, professional purpose, and social value when technological change displaces or devalues human abilities [25]. Across theoretical perspectives, this erosion takes different forms: Social Identity Theory highlights how AI displacement can threaten the prestige or relevance of professional communities (e.g., journalists, designers, or legal professionals), weakening both personal and collective identity (see [6] for more discussion of this). Role Identity Theory [24] emphasises that when AI automates role-defining tasks, people may lose clarity about their place in society, may become undervalued or invisible, especially if human contributions are perceived as slower, costlier, or less accurate than AI outputs [5]. This is consistent with research which demonstrates that self-identity is an important determinant of self-esteem (e.g. [23]).

Future Identity in an AI Age

AI's influence on human identity will not be limited to isolated professions or digital platforms. Its continued development will reshape how individuals and societies define identity in at least four important ways.[1]

- **Shifts in work and professional identity.** Many professions will be redefined as AI automates routine and even creative tasks. For some, this will mean loss of status or role clarity. For others, new hybrid roles built around human–AI collaboration may emerge, emphasising oversight, creativity, and interpersonal skills. Designing these roles carefully is essential to ensure that AI augments rather than erodes human agency and skill application [18].
- **Rise of digital identity ecosystems.** The growing use of digital identity systems, including decentralised or self-sovereign identity models, will change how people prove, assert and manage "who they are". These systems may empower individuals with greater control but risk fragmenting identity across multiple platforms, organisations, and jurisdictions.
- **Cultural narratives and societal values.** The stories societies tell about AI will strongly influence whether people view it as a threat to human uniqueness or as a tool for human empowerment. Narratives of replacement and obsolescence can accelerate identity erosion, while narratives of augmentation and resilience can sustain confidence in human agency and worth.
- **Well-being and psychological adaptation.** The psychological consequences of AI adoption will extend beyond productivity or efficiency. Individuals may experience technostress, anxiety, or loss of self-worth if they feel displaced, something which can shaped by the culture and practices of the organisation [16].

[1] It is of course also true that humans shape technologies (see for example [6,18]). However here we are concerned about the of a particular impact on humans of technology – identity erosion.

Conversely, AI could enable identity growth by creating opportunities for reskilling, new forms of creativity, and novel ways of expressing oneself in digital environments.

- **Human-AI Blurring:** As AI agents become more autonomous and human-like, distinguishing between human and machine actors will become more difficult. Also, the flipside of augmentation is uncertainty about what it is to be human. This in turn raises ethical and legal questions about personhood and accountability [20].

Critical Issues and Opportunities for TIPSS

Preserving identity in an AI-driven world requires more than technical safeguards: it demands design choices, policies, and cultural narratives that affirm the enduring value of human agency. The challenge ahead is ensuring that technological progress strengthens rather than diminishes the identities through which individuals and communities find meaning and purpose. Each pillar of the TIPSS framework presents both immediate risks and emerging solutions, though their implementation faces significant practical barriers.

Trust - Rapid AI integration into professional and personal life can bypass informed consent for identity-related data processing. This loss of control over one's identity and digital autonomy may trigger resistance to technology adoption as individuals seek to reclaim agency [16]. Without transparent AI decision-making and clear human–AI collaboration boundaries, trust deficits compound across all other TIPSS dimensions.

To address these trust and consent challenges, emerging solutions include dynamic risk scoring using behavioural biometrics (i.e., systems that continuously monitor usage patterns and habits to verify identity in real time). Decentralised trust infrastructures using blockchain can distribute verification across stakeholders, while maintaining immutable audit trails [22]. Explainable AI techniques can expose decision pathways and justifications to users and regulators, strengthening transparency and accountability [27].

However, these approaches face substantial implementation challenges, requiring extensive user training data, remaining energy-intensive and technically complex for mainstream adoption. Moreover, lack of incentives may limit the voluntary adoption of transparent alternatives without regulatory pressure.
Identity - AI systems increasingly display human-like attributes such as authorship, voice, and decision-making authority, challenging how people construct and sustain their identities. This shift threatens not only personal data but the social and psychological value of being recognised as an autonomous, competent, and creative agent.

Decentralised Identifier (DID) frameworks offer individuals cryptographic control over identity attributes via selective disclosure, allowing them to reveal only necessary information for specific contexts [11]. Personhood credentials (PHCs) can attest to human authorship and accountability without exposing unnecessary personal data [4]. Context-aware identity models that scope claims and permissions to specific situations, limiting misattribution.

However, DID systems currently lack interoperability standards, creating fragmented identity ecosystems. PHCs need to be designed to be robust against changes in a person's circumstances and face issues about equitability of access: if providers use personhood credentials to limit access to particular services, some groups without PHCs may be systematically excluded [1].

Privacy - AI systems that model human behaviour, voice, or writing styles, can inadvertently expose personal traits or preferences. Even without explicit identifiers, AI can reconstruct an individual's identity from behavioural or biometric patterns, creating new vectors for identity theft. This makes self-sovereign identity, which is "the idea of allowing individuals to control the identifying information they provide" [8], increasingly critical.

Privacy-preserving machine learning techniques such as federated learning with secure aggregation and differential privacy can train global models while keeping raw data in its original location. Consent-aware data architectures can provide granular, revocable permissions with clear provenance tracking, enforcing purpose limitation and reducing exposure. High-fidelity synthetic data, when properly validated for utility and leakage, can augment or replace training sets to minimise re-identification risk.

These privacy-preserving approaches, however, involve significant performance trade-offs limiting model accuracy as privacy guarantees strengthen. Federated learning can outperform isolated local models, while it imposes coordination overhead and may perform poorly with heterogeneous data distributions.

Security - Identity spoofing and deepfakes can enable highly personalised AI-driven social engineering attacks.

When AI embeds within identity management systems, security breaches can allow attackers high-level access to sensitive services and resources, potentially leading to financial loss, reputational damage, and compromised safety.

Identity-first security frameworks, adopting Zero Trust principles, continuously verify user, device, and context for every access request. Combining least-privilege access controls with micro-segmentation and AI-enhanced threat detection can identify deepfake-driven account takeovers, lateral movement, and credential misuse. These systems can enable automated containment and risk-adaptive access controls [15].

Nevertheless, in addition to infrastructure costs required to implement these systems, the arms race between AI-powered attacks and defences means that security measures face fast obsolescence, requiring continuous investment and updates.

Safety - Identity erosion can produce severe psychological impacts, including anxiety, loss of self-worth, and disengagement, particularly in professions under heavy AI adoption pressure. At scale, displacement of identity-linked roles risks destabilising communities and entire industries, creating social tensions that extend beyond individual wellbeing.

Regulatory frameworks lag significantly behind technological development. Organisations face competitive pressures that make comprehensive safety measures appear as unnecessary overhead, while the distributed nature of AI deployment makes consistent safety standards difficult to implement and monitor across diverse contexts and jurisdictions.

Human-centred AI design can preserve agency and dignity through ethical digital tools that ensure opt-in consent, provide clear explanations, and incorporate human-in-the-loop controls. These design principles require reinforcement through regulatory frameworks mandating algorithmic accountability via impact assessments, independent audits, incident reporting, and accessible redress mechanisms [27].

Summary

The path forward requires acknowledging that technical solutions alone cannot

address TIPSS challenges. Success depends on aligning economic incentives with human welfare, developing adaptive regulatory frameworks, and nurturing human agency and creativity alongside technological advancement.

References

[1] Adler S, Hitzig Z, Jain S, Brewer C, Srivastava V, Christian B, Trask A. Personhood credentials: artificial intelligence and the value of privacy-preserving tools to distinguish who is real online. 2024.

[2] Baumeister RF. Identity, self-concept, and self-esteem: The self, lost and found. In: Hogan R, Johnson J, Briggs S, editors. *Handbook of Personality Psychology*. Academic Press; 1997. p.681–710. Available from: https://doi.org/10.1016/B978-012134645-4/50027-5.

[3] Books A, Goffman E. The presentation of self in everyday life. London: Allen Lane. 1969.

[4] Brunton F, Nissenbaum H. *Obfuscation: A User's Guide for Privacy and Protest*. MIT Press; 2020.

[5] Brynjolfsson E. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. Vol. 236. WW Norton Company; 2014.

[6] Chuah E, Reid I, McAlaney J, Vigano L, Jacobs N, Winter P, Dover R, Lu Y. (2026) Artificial intelligence and the future of human employment. *SPRITE+ Explainer* #004. 2026. Available from: https://77265d30-ac80-43a7-a339-19564066f602.usrfiles.com/ugd/77265d_17f38282ccd340ee9128090ba1cfc0c0.pdf

[7] Coeckelbergh M. *Using Words and Things: Language and Philosophy of Technology*. 1st ed. Routledge; 2017. https://doi.org/10.4324/9781315528571.

[8] Elliot M, Mandalari AM, Mourby M, O'Hara K. *Dictionary of Privacy, Data Protection and Information Security*. Edward Elgar Publishing; 2024.

[9] GOV.UK. Data Ethics Framework: glossary and methodology. 2020. Available from: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-glossary-and-methodology.

[10] Hogan B. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bull Sci Technol Soc*. 2010;30(6):377–86. https://doi.org/10.1177/0270467610385893.

[11] Huang K, et al. Zero-Trust Identity Framework for Agentic AI. 2023. Available from: https://arxiv.org/abs/2306.00953.

[12] Jones E, Dunlop C, Ghani B. What is a foundation model? 2023. Available from: https://www.adalovelaceinstitute.org/resource/foundation-models-explainer.

[13] McAdams DP. The psychology of life stories. *Rev Gen Psychol*. 2001;5(2):100–22. https://doi.org/10.1037/1089-2680.5.2.100.

[14] Mead GH. *Mind, Self, and Society*. Chicago: University of Chicago Press; 1934.

[15] Microsoft. Digital Defense Report. 2023. Available from: https://www.microsoft.com/en-us/security/business/security-intelligence-report.

[16] Mirbabaie M, Brünker F, Möllmann Frick NRJ, et al. The rise of artificial intelligence – understanding the AI identity threat at the workplace. *Electron Markets*. 2022; 32:73–99. https://doi.org/10.1007/s12525-021-00496-x.

[17] Papacharissi Z, editor. *A Networked Self: Identity, Community, and Culture on Social Network Sites*. 1st ed. Routledge; 2010. https://doi.org/10.4324/9780203876657.

[18] Parker SK, Grote G. Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. *Appl Psychol*. 2022;71(4):1171–204. https://doi.org/10.1111/apps.12241.

[19] Pinch TJ, Bijker WE. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Soc Stud Sci*. 1984;14(3):399–441.

[20] Puzio A. AI and the disruption of personhood. In: *AI and Ethics*. Springer; 2023.

[21] Snow DA, Anderson L. Identity work among the homeless: The verbal construction and avowal of personal identities. *Am J Sociol*. 1987;92(6):1336–71. https://doi.org/10.1086/228668.

[22] Stanford Journal of Blockchain Law & Policy. Decentralized identity and democratic integrity. 2023. Available from: https://stanford-jblp.pubpub.org.

[23] Stets J, Burke P. Self-esteem and identities. *Sociol Perspect*. 2014; 57:409–33. https://doi.org/10.1177/073112141453614.

[24] Stryker S, Burke PJ. The past, present, and future of an identity theory. *Soc Psychol Q*. 2000;63(4):284–97. https://doi.org/10.2307/2695840.

[25] Susskind D. *A world without work: Technology, automation and how we should respond*. Penguin UK; 2020.

[26] Tajfel H, Turner J, Austin WG, Worchel S. An integrative theory of intergroup conflict. In: *Intergroup Relations: Essential Readings*. 2001. p.94–109.

[27] West SM. The emotional politics of AI: Introspection and identity in the age of algorithms. *AI Soc*. 2023.

[28] Zuboff S. *The Age of Surveillance Capitalism*. PublicAffairs; 2019.