



DIGITAL
Enlightenment
STUDIES

VOLTAIRE FOUNDATION
UNIVERSITY OF OXFORD

Computational hermeneutics: why study canonical figures like David Hume in the age of AI?

Mikko Tolonen, University of Helsinki, mikko.tolonen@helsinki.fi

Tolonen M. 2026. 'Computational hermeneutics: why study canonical figures like David Hume in the age of AI?.'

In: *Digital Enlightenment Studies* 3, 1–21.

DOI: <https://doi.org/10.61147/des.64>

This article argues that figures like David Hume (1711–1776) should be studied in the age of AI not despite computation but through it. Drawing on Enlightenment studies and computational history, I sketch a triadic model of the book – as artefact, data object, and vehicle of meaning – and show how metadata, text and image reuse, translation mining, and model-assisted workflows let us trace reception, canon formation, and semantic change at scale. However much we may wish to move beyond canonical figures, the canon remains one of the main structures through which visibility, transmission, and exclusion are historically organised; understanding how it was made is often the most effective route to recovering those figures left outside it. The aim is not to automate interpretation, but to relocate it: canonical authors become testing-grounds for reproducible, data-aware intellectual history with broader relevance in an AI-saturated world.

Keywords: AI, canon, computational history, David Hume



A preliminary version of this article was given as the Voltaire Foundation Lecture in Digital Enlightenment Studies, at the Bodleian Library, Oxford, 9 May 2025.¹

1. Introduction

The conventional wisdom about computation in the humanities has long involved a methodological compromise: algorithms detect patterns, then the machine is set aside and the real interpretation begins. This division of labor – often described as ‘mixed methods’ (Prescott 2023) – once made sense. It lowered the threshold for meaningful use of algorithms while protecting the prerogatives of close reading. But it also caps ambition. If computation is always a prelude, we miss a larger opportunity: a genuinely iterative practice in which models are shaped by humanities questions, data work is planned around historiographical insight, and interpretation happens with the machine and the scholar both in the loop.

This essay argues for that reframing from the perspective of Enlightenment studies and computational history. It draws on over a decade of work that moves from bibliographic description to explanatory models of canon formation; from printers’ ornaments to semantic similarity; from union catalogues to multilingual alignment; and from solitary reading to shared infrastructures.

At the center lies a question I am asked with disarming regularity: in the age of AI and high-performance computing, why spend time on 18th-century figures like David Hume? Shouldn’t the supercomputers be applied to something more urgent – more contemporary, more useful?

My answer is no. Studying Hume through computational hermeneutics is not a retreat from relevance; it is a way of rethinking how meaning circulates, how intellectual systems evolve, and how cultural authority is constructed. It is as urgent – and contemporary – as any challenge in the humanities and social sciences today.

The first reason is economic: Hume is worth understanding. Knowledge for its own sake is not an academic luxury – it is a core component of the civic and long-term orientation that underpins modern societies. Distilled through figures like Keynes and Hayek, this foundation has often been cast not as a rival to economic efficiency but as its precondition: two sides of the same coin. The risk today is not that we flip that coin, but that we toss it aside. No cryptocurrency can replace civility.

Hume stands at the center of that inheritance. If we reduce civil society to market logic alone, we risk eroding the trust, institutional coherence, and cultural literacy that markets themselves depend on. Hume’s

¹ I am grateful to Nicholas Cronk, Gillian Pink, and the entire team at the Voltaire Foundation for their warm hospitality and longstanding support. I also thank the two anonymous reviewers for their valuable feedback, and Gillian Pink and Alison Oliver at *Digital Enlightenment Studies* for their careful editorial work. The alternation between ‘I’ and ‘we’ in this paper reflects the distinction between my individual perspective and the collaborative nature of the research carried out within the Helsinki Computational History (COMHIS) group. Rather than listing all collaborators by name, I have aimed to reference the most relevant co-authored publications throughout the paper. These citations, of course, express more than gratitude – they are the foundation of the argument.



or Adam Smith's political economy was never about the mechanics of trade alone, but about the moral and institutional conditions that make exchange sustainable in the first place (see Rasmussen 2005). Without space for humanities to thrive – and without a viable path for careers built on such knowledge – we sever the roots that nourish our economic systems. When short-term efficiency undermines the ecosystem that produces long-term insight, the collapse is slower but ultimately deeper than any market crash.

Hume scholarship rests on a foundation that recognizes him as a multifaceted author – not only a philosopher, but also a historian (Spencer 2013), essayist (Harris 2015, pp.143–97), political economist (Hont 2005), and theorist of civil society (Robertson 2009, pp.256–324). None of the computational work underlying this essay would have been possible without years of working alongside scholars who understand this – discussing sources, refining questions, and arguing about interpretations. Their insights are part of a broader understanding shaped by a traditional scholarly community – one sustained by institutions and funding bodies that have protected academic freedom, including the obligation to confront the more troubling dimensions of Hume's legacy. 18th-century wealth was deeply entangled with slavery and colonial extraction, exemplified by the sugar trade and its role in British elite life (see Kidd 2025). Dealing with these entanglements is part of the scholarly task – studying the past as it was, not as we wish it had been.

The point I want to stress is that the current 'understanding of Hume' has been the product of decades – indeed, lifetimes – of scholarly commitment (see Garrett 2025). Foundational work has emerged through laborious efforts to read the 18th century in its full complexity. But that kind of devotion, which not all scholars can afford, has also shaped the field's development in uneven ways – leaving certain perspectives un(der)explored, and delaying progress that might have occurred more rapidly under more collaborative or structurally supported conditions. For example, when 20th-century philosophers embraced Hume primarily as a theorist of metaphysics and epistemology, his contributions as a historian were sidelined for years.²

The nature of humanities scholarship – its deep silos, closed networks, slow maturation, and reliance on heroic individual effort – is itself part of the story. By contrast, the core claim of digital humanities is that accelerating research through computational methods, interdisciplinary collaboration, and new methodological frameworks – combined with more cumulative and reproducible forms of data and knowledge – can both correct embedded biases and enable deeper, more systematic forms of interpretation. That is the promise of computational hermeneutics.

The second part of my answer is methodological. Hume brings into focus a broader challenge: how to expand and scale the contextual insights of intellectual history. One helpful concept here comes from

² When I attended my first Hume Conference in the early 2000s, submissions were grouped into three categories: A for epistemology and metaphysics, B for moral philosophy, and C for everything else. The center of gravity was unmistakably clear. Since then, the Hume Society has made substantial efforts to foster greater inclusivity and to broaden the field's scope – without reinforcing a hierarchy between so-called 'hard' and 'soft' areas of scholarship.



the social sciences: the cultural constituent – a practice, belief, or value that becomes embedded in culture and shapes meaning and behavior independently of any one individual’s intention. Canonical authors operate in this way too. Their ideas do not merely circulate; they are absorbed, reused, and rephrased, taking on meanings that detach from authorial intent. Once this happens, texts begin to move differently – reframed for new publics, embedded in new genres, and transformed by shifting norms.

Intellectual history has traditionally focused on reconstructing authorial intent, asking for example what Hume was doing when he added a racist footnote (Garrett and Sebastiani 2017) in one of his essays (e.g. Skinner 1969; see also Pocock 1987). Yet this task must be complemented by attention to reception – the ways texts are appropriated, reinterpreted, and repurposed across time (Thompson 1993; Melve 2006; Hanisch 2007; and especially Jauss 1974). While we may grasp that meaning depends on context, defining ‘context’ itself remains elusive.³ We need to think about how texts become effective. Reception studies is not simply about tracing the afterlife of a text – often dismissed as secondary – but it ought to be a central concern in early modern thought. Serious engagement with intellectual history requires seeing texts not just as responses to specific predecessors but as interventions in complex interpretive ecologies shaped by cultural constituents among other forms of ideas (see LaCapra 1980).

A common misconception underlies this: that meaning in a printed book – or even in a single footnote – is stable. But meaning shifts with every change in material context, editorial mediation and audience (Ricoeur 1971). Authorial intention, in the absence of direct testimony, is never fully recoverable. The idea here is not to endorse any romantic notions about the ‘death of the author’ (Barthes 1977). To understand the development of ideas, we need to account for authors and other agents with clearly-defined roles in meaning-making. Without them, we are left with an unstructured mass – mere frequencies of words and textual patterns. Tracing how language evolves is important, but it is not sufficient for serious intellectual history.

Consider how Hume’s remarks on race and slavery were deployed in the late 18th and 19th centuries: they served *both* abolitionist and pro-slavery arguments, illustrating how meaning is forged in contestation (Spencer and Tolonen 2025). Today, Hume’s statue stands on Edinburgh’s Royal Mile, while his writings face justified criticism. Efforts to remove or rename in response to these legacies reflect current moral and political frameworks as much as historical reckoning. They do not erase Hume; they redefine him. Cultural constituents persist not by remaining fixed, but by accruing new – and often contradictory – meanings. At the same time, as we account for such shifts in meaning, a more difficult but necessary question emerges: to what extent might later interpretations of Hume’s views on race be situated in relation to, or even shaped by, the epistemological structures his works helped establish?

³ For a related discussion about context with respect to concepts, see Kuukkanen (2008); and Kaukua and Lähtenmäki (2020). See also Koikkalainen (2011).



To label these shifts as mere anachronism and focus, rather, on ‘what Hume really said’ is to misunderstand the interpretive life of ideas. Even within authors’ lifetimes, texts evolve in ways they might not have anticipated or endorsed. Quite often, too, the author’s own understanding of the meaning of his or her earlier work mutates. Historical (mis)alignments are not simply errors; they are part of how meaning is made. Tracing how passages are republished, reframed, misread, and edited over time demands tools that go beyond traditional close reading. This is where the contextualist emphasis of the Cambridge School provides a foundation – though it was not originally focused on remediation, which in hindsight is understandable given the complexity of the task (Skinner 1966). It is precisely here that computational methods become necessary (Tolonen and Ryan 2026).

To understand Hume, we must trace not only how we think he read others, but how those others were understood in his time, how those readings changed, and how Hume’s work contributed to – and was shaped by – these evolving contexts. Canonical figures like Hobbes or Mandeville were mediated through interpretive filters and ideological frames before Hume encountered them (see Tolonen 2010, pp.223–59). Meaning is shaped not just by authors but by chains of reception already in motion. And once we extend the frame beyond canonical figures, the network expands. Traditional intellectual history lacks tools to analyze these dynamics at scale. But that is now changing (Tolonen 2026).

With data science, we can begin to follow how meaning circulates and mutates – across languages, across publics, across time. We can measure not just where influence existed, but how it changed. Reception, in this sense, becomes a methodological opportunity: not a flattening of interpretation, but its deepening. Enlightenment figures like Hume and Voltaire matter not because they encapsulate their era’s thought, but because they sit at the nexus of institutional authority and public uptake. Their significance becomes legible only in relation to the translators, compilers, pamphleteers, and lesser-known figures who surround them (Emerson 2008).

Reception is not a postscript to interpretation – it is its condition. Hume’s essays, like any influential work, cannot be understood apart from the formats, prices, publishers, and networks that shaped their circulation. Book history matters and intellectual history must operate at scale beyond just reading the text. We can trace how key concepts like commerce, power, or corruption travel across 18th-century discourse, testing how they align with or challenge major contemporary historiographical interpretations (Tiihonen and Tolonen 2026). That task requires computational hermeneutics – text mining, multilingual alignment, network analysis – not as optional extras but as essential methods. Without them, we risk reducing reception to anecdote, and overlooking how meaning is actually made.

The Enlightenment – with its historical distance and dense textual record – offers a rare opportunity to study these processes in motion. Studying Hume becomes not just an act of interpretation, but a way to model how ideas evolve and intellectual systems change. If we want to understand how meaning circulates, reshapes, and becomes embedded in culture, there are few better places to begin.



2. Bibliographic data science

An important part of contextualizing Hume is recognizing him as a participant in early modern public discourse. One productive way to approach this is by imagining that discourse as a vast network, composed of diverse actors, features, and layers. To navigate this complexity, we ought to adopt a triadic model of the book⁴ – understood simultaneously as a physical artefact, a data object, and a vehicle of meaning. The first layer foregrounds production, format, and material circulation; the second treats books as structured and unstructured data amenable to modeling; the third encompasses interpretation – claims, arguments, genres, and the ways in which readers transform them. Traditional intellectual history privileges the third; book history the first; computer science the second. The promise of computational history is not that one layer eliminates the others, but that all three belong to a single analysis, because meaning in the past is inseparable from the mechanisms that made it legible, durable, and mobile. In the context of Enlightenment studies, this means reading Hume and, at the same time, reading the catalogues and imprints that show how his works were mediated; it means reconstructing publishing ecosystems and, at the same time, modeling semantic change in the books and newspapers that reviewed, excerpted, praised, and condemned him.

The obvious obstacle to such work is that the evidentiary foundations in the humanities were never designed – let alone systematically assembled – to behave like empirical datasets ready for straightforward measurement. National bibliographies or union catalogues – like the English Short Title Catalogue (ESTC) – and commercial full-text corpora such as Eighteenth Century Collections Online (ECCO) are indispensable, but their metadata are uneven, the edition information uncertain, their coverage patchy, their OCR variable and their licensing inconsistent (Tolonen, Mäkelä and Lahti 2022; Mäkelä et al. 2025). I have spent a decade with my research group, Helsinki Computational History group (COMHIS), dealing with these issues. It is at this point that our methodological stance departs from both the cautious abstention typical of traditional intellectual history and the uncritical enthusiasm often found in digital humanities. Rather than treating these issues with data as a reason to stay in the realm of the analog – or as noise that can be ignored – we test what the data can bear and we build what is missing, possible, and necessary. That has meant harmonizing and linking metadata across catalogues and allied sources; developing principled procedures for determining editions; finding ways to deal with uncertainty when we model coverage and price; associating catalogue records with page images and full text; and, crucially, producing workflows and interfaces through which scholars can study, annotate, and correct the signals. We have called this *bibliographic data science*: the systematic use of library catalogues and related corpora as structured evidence for cultural and intellectual history, subject to the same norms of clarity, error accounting, and hypothesis testing that we would expect in any serious scholarly endeavor (Tolonen et al. 2019).

⁴ The concept of ‘the book’ here includes also manuscripts.



This approach did not arise from theory or examples from natural science. It grew out of very practical beginnings: close reading in archives, the reconstruction of arguments in the writings of Mandeville and Hume, studying their publishing histories and the perennial questions of Enlightenment literature – greed, self-esteem, skepticism, commercial society, and the uneasy virtues of luxury (Tolonen 2010). Frustration in my analog days was structural rather than intellectual. With the tools then available, it was possible to imagine what studying cultural constituents would look like, but only to follow a handful of citation trails. It was not possible to model how concepts travelled across genres, languages, and publics at a scale that would make explanatory claims robust. Early digital resources were behind paywalls, on CD-ROM, or brittle to use, OCR was uneven, metadata inconsistent, and books and newspapers lived in separate silos. That combination forced a change of method: away from solitary reading supplemented by ad hoc queries of the ESTC, toward collaborative infrastructure that would make reasoning with evidence repeatable, inspectable, and extendable.

Autobiographically, this line of research meant assembling an interdisciplinary team one step at a time – historians, computer scientists, linguists, and librarians – and agreeing on a compact: research questions come first; data and models serve the questions; and every step must be explicit enough that another scholar can rerun it, disagree with it, and see what changes. The result is not the abandonment of traditional intellectual history, but a regrounding of it. Books remain artefacts, data objects, and vehicles of meaning; the difference is that the evidentiary trail linking those layers is now engineered rather than improvised.

In computational history in Helsinki, our first scholarly move in this direction was descriptive. We asked what could be counted reliably and meaningfully across early modern print culture with respect to canon formation. These early studies charted formats, page counts, places, and languages from 1500 to 1800 across northern Europe, asking, for example, how and when vernaculars displaced Latin (Tolonen et al. 2025; Wu et al. 2026). What we found would not fit the easy narrative that ‘the vernacular rises’. Latin held on longer in university towns; commercial centers switched to vernacular sooner; and genres behaved differently – dissertations were not pamphlets, and theology evolved at its own tempo. These simple maps matter not because counting is good in itself, but because you need a defensible picture before you can explain anything.

Two practical lessons followed and have guided everything since. First: the edition matters. You have to group printings into work-level families, otherwise ‘reprint’ and ‘new edition’ blur and your growth curves are often just catalogue noise (Ijaz et al. 2019). Second: ‘good enough’ does not necessarily mean ‘sloppy’. It means setting thresholds, designing analysis that tolerates a certain amount of error in the source data, and making assumptions explicit so they can be rerun and adjusted.

From description we moved to explanation and to questions of greater import for traditional intellectual history. One line of enquiry on canon formation asked whether the 18th century’s reprint boom broadened or narrowed cultural foundations (Tolonen et al. 2021). The answer was in some sense expected, but still



striking. Reprints of those works that became canonical increased in number, while the diversity of authors shrank. In classical reception, the effect is particularly sharp: the presence of Ancient authors in Britain grew in absolute terms, while contracting to a narrower subset (Fantoli et al. 2025). This empirical Matthew effect (where visibility, influence, and resources accrue disproportionately to those who already have them) does not settle questions of value, but it changes how we talk about them. It shows that the emergence of ‘the classics’ that Hume, among others, inherited was not a neutral accumulation of learning; it was a selective amplification that privileged some trajectories and silenced others, mediated by publishers, markets, and pedagogical regimes. A related line of research modeled book price probabilities to revisit the persistent myth that the 1774 copyright ruling suddenly democratized reading. The data suggest otherwise. Even as print expanded, affordability constraints remained stubbornly constant. Reading became more widespread (especially when upper-class women became a crucial new readership), but it remained a privilege (Tiihonen, Lahti and Tolonen 2024). The conclusion does not deny democratization; it corrects mechanism and timeline, and places publishers and price structures at the center of explanation. It also makes it possible to understand why it was so important for Hume to disseminate his essays in newspapers (Spencer and Tolonen 2025).

A third line of scholarship on bibliographic data foregrounded the role of place and network. Reconstructing Scottish Enlightenment publishing with network methods revealed distinct ecosystem roles in Edinburgh and London. We began by placing Hume’s move from Edinburgh to London in its full publishing context and by making earlier book history findings operational in code (Sher 2006). University printers, commercial houses, and hybrid shops mediated genres differently and shaped the flow of ideas in ways that social-intellectual histories ought to take seriously. The ‘careers’ of authors and publishers emerge not as static biographies but as trajectories through relational space – as shifts in centrality, brokerage, and role (Ryan and Tolonen 2024). If publishers are filters – selecting, financing, sequencing, bundling – they explain not just what was published but how certain discourses consolidated as ‘the Enlightenment’ for later readers. The point is not to deny authorial agency; it is to nuance it with other mechanisms that make agency legible.

The approach to structure and content also co-evolved in our strategy over the years. Unstructured text and images were not only objects of analysis; they became sources for new structured data. Sermons, pamphlets, or page images could be turned into structured lists of themes, people, places, or typographic features. For example, one of our later projects has been the extraction and clustering of printers’ ornaments – hundreds of thousands of headpiece variants grouped by exact visual similarity and tracked over time. What had once demanded scissors and paste (e.g. Maslen 1973) could now be approached as a large-scale attribution problem in which Dublin, London, and Edinburgh styles diverged, house practices left signature reuse patterns, and ornament trajectories after a printer’s death supplied evidence for continuity and ghost imprints (Wang et al. forthcoming). Image segmentation and layout detection are not peripheral matters



or a separate technical hobby. Of course, others have also done excellent work on extracting and clustering woodcut ornaments and we build on that (Wilkinson, Briggs and Gorissen 2021; Abhishek, Bergel and Zisserman 2021). What distinguishes our approach is that our primary interest is not to create databases, but rather to develop a holistic approach toward an ever-growing data ecosystem. We integrate these new pieces of evidence into a single bibliographic data science workflow – tied to harmonized metadata, reuse detection, and page image verification. In this context, printers’ ornaments sit alongside paratexts, formats, prices, and publisher/printer networks as analyzable variables. This approach is a necessary foundation for treating physical artefacts as evidence at scale and it reopens attribution questions that matter for both book and intellectual history.

3. Reception at scale

Meanwhile, reception studies pushed us from catalogue level relations to latent ones.⁵ How do ideas travel when there is no explicit citation? One answer begins with literal overlap. Using the IT Center for Science in Finland (CSC) supercomputers and large-scale text reuse detection on 18th-century corpora, we identified hundreds of thousands of near verbatim links between Hume’s *History of England* and earlier historiography, then qualified, clustered, and compared those matches against claims about his political leanings and source use. What makes this useful is not only the volume of candidates but the way they can be inspected (Vaara and Tolonen 2025). After conducting a few similar case studies, it became clear that this could form the basis of a useful end-user interface for historians. Through the Reception Reader, a historian can now hover to see who quotes whom and where, click through to page images and start reading (Rosson et al. 2023). The tool simply exposes patterns, quickly and at scale, for human inspection. With that workflow, reception becomes observable across hundreds of thousands of volumes, millions of pages and roughly half a billion text reuse pairs (Mahadevan et al. 2025). In the case of Hume’s *History*, this meant testing long-standing assumptions about Royalist or Whig dependency by showing, chapter by chapter, which authors were reused, in what proportions and when.

What soon became evident is that the same mechanism is exactly what an editor needs to find changes between editions. As a consequence, a collaborative scholarly edition of Hume’s *History of England* (forthcoming, OUP), is using the pipeline that surfaces textual overlaps across early modern corpora and applying it to effect an edition-to-edition comparison for a single work. Once every printing has been digitized, segmented, and converted to text using OCR with layout preserved (Vesalainen, Tolonen and Ruotsalainen 2024), the reuse engine establishes a baseline of sameness – what is carried forward verbatim – and directs attention onto the remainder: the omissions, additions, relocations, and rewritings that constitute the work’s evolving form. Eventually the interface will become a scalable critical

⁵ On integrating reception into historical explanation, see Thompson (1993).



editing tool: editors move between global comparison and page images. Without this kind of end-to-end methodology – document layout, accurate OCR, overlap detection, and text to image verification – producing a reliable edition of Hume’s *History* would remain the century-long, largely manual ordeal that earlier generations chose not to undertake. With it, editorial judgment is not replaced but concentrated: the machine finds the stable bedrock and suggests changes between editions; the editor decides what the changes actually mean and the process continues iteratively through a human–machine dialogue.

Because this approach is modular – image handling and layout analysis, recognition, overlap/gap logic, and an editor-facing interface – the payoff is not confined to Hume. It also enables us to move past the usual handful of canonical figures and into the wide middle of public discourse – the large group of near-canonical authors whose works shaped local and continental debates but have received far less attention, often only because the labor of editing them has never seemed likely to pay off. If the tools work for Hume’s *History*, they will also work for William Robertson – and so on. They also complement the other end of the spectrum, where canonical giants already have their critical editions: here, our methods shorten the road for supplements, errata, or born-digital companions by making cross-edition control affordable and auditable. In that sense, the Hume’s *History* as critical edition project is both a proof of concept and a scalable template. It shows how reception tooling becomes editorial tooling; how a similar interface lets a historian chase a quotation across newspapers and an editor trace a paragraph across editions; and how a single, reusable workflow can serve the field’s long-standing need to move beyond a handful of painstakingly edited monuments toward a broader, more equitable ecology of reliable texts.

Alongside that edition-facing work, we have also ventured into grounded studies on Hume’s prose, beginning with Scotticisms in his works compared to the full ECCO corpus (Tiihonen et al. 2023). Large digitized corpora let us move beyond anecdote and other evidence in authors’ letters: we assemble historically attested lexical features and then query, for example, Hume’s essays at scale while benchmarking against period baselines drawn from comparable 18th-century English prose. The result is a map – by essay, register, and local context – of where Scotticisms actually occur and how frequent they are relative to contemporary norms tested against Hume’s list of Scotticisms. No claims about intention are required; the point is to turn scattered observations into testable distributions, to see whether suspected Scotticisms cluster around particular topics or genres and to quantify rarity versus routine. This kind of data-aware study belongs in the same toolkit as our reception and edition pipelines: it shows that the same computational discipline that helps trace paragraphs across editions can also illuminate the fine texture of an author’s language and questions that have surprising intellectual weight, such as that of the occurrence of Scotticisms (Tolonen et al. forthcoming 2026).

4. Translation mining and meaning matching

Much of Enlightenment discourse moves through paraphrase, selective translation, the incorporation and reconfiguration of existing ideologies, and polemical inversion – forms that rarely leave neat bibliographic



fingerprints. Making these pathways observable means working at the level of meaning, not just surface strings. An example of this in our work is what we call ‘translation mining’: it treats books as sequences of semantically comparable segments, then aligns them across languages to expose corridors of reuse that run from faithful rendering to hostile recoding. In practice, the pipeline is simple to describe – segment, embed, align, and rank – but the consequence is profound: once the semantic spaces of, say, English and French are mapped onto one another, comparative intellectual history becomes operational rather than aspirational. We no longer infer ‘influence’ from hints and intuition alone; we can test it, visualize it, and argue about gradations instead of binaries.

Once meaning is made computable in this way, the field can move beyond anecdotes toward systematic typologies. Our efforts on translation mining support a data-driven taxonomy of practices – from single-volume renderings and volume reconfigurations to collections, omissions/insertions and single- or multi-fragment borrowings (Hinderks et al. forthcoming). Visually, these appear within similarity heatmaps as distinctive ‘diagonals’ and breaks; analytically, they correspond to choices translators, editors, and publishers made in response to markets, censorship, pedagogy, and politics. The upshot is a more granular and accurate vocabulary for cultural transfer: faithful translation, selective appropriation, hostile inversion, pedagogical paraphrase, and the many hybrids in between. Crucially, these categories emerge inductively from the evidence rather than being imposed in advance, which is why they travel well across genres – sermons, pamphlets, encyclopedias, essays – and across the canonical/non-canonical divide.

What matters here is scale. A large corpus should not be treated merely as a finding aid that helps us to locate a few interesting cases. The real aim is to ask population-level questions that were previously painstakingly difficult to answer: What share of Hume’s arguments circulated translingually, and through which channels (books, gazettes, anthologies)? How often are his core claims preserved, and how often reframed, and in which moral or political registers? Do translation corridors cluster by genre or decade, and do they align with shifts in publishing geography? At the macro level we can measure the ‘interpretive entropy’ of a given essay – how many distinct semantic trajectories it takes on over time – and compare those distributions across authors and traditions. At the meso level we can model selection pressures: which parts of Hume are persistently excerpted or paraphrased (and which fall into obscurity), across which publics. Zooming in remains essential for assessing the pertinence of specific alignments, but the target has changed: the goal is to recover Hume’s 18th-century footprint as a whole – his actual presence as a multilingual, multi-genre phenomenon, as a cultural constituent – rather than a mere handful of emblematic receptions. In other words, translation mining lets us move from ‘examples that illustrate a thesis’ to ‘measurements that define what the thesis will be’.

The shift here is not just one of scale; it is also epistemic. Classical text reuse catches verbatim overlap. By contrast, meaning-level matching – powered by multilingual sentence embeddings and strengthened by large language models (LLMs) – recovers partial translations, layered paraphrases, and composite works



that stitch together several sources.⁶ The machine does not decide what counts as engagement; it proposes candidates that a historian can accept, refine, or reject. That feedback – thresholds, false positives, edge cases – loops back into the retrieval settings and the heuristics for classifying types of transfer. In this sense, AI becomes a hermeneutic partner. It multiplies the number of logical places where interpretation can begin or where it goes next, and it does so across language boundaries that used to dictate separate historiographies.

Quotation is only the visible tip of transmission. Laura Nicolì's research on the French reception of Hume's *Essays* shows how far traditional scholarship can go: by following catalogues, paratexts, and periodical runs, she reconstructs the relays that carried Hume across the Channel – rival translations of *Political Discourses* in 1754, essays excerpted and retranslated in gazettes, with the *Journal étranger* as a key node, and even the *Mercure de France* printing 'Of the middle station of life' after Hume suppressed it in English (Nicolì 2025). None of this necessarily requires AI; it is all recoverable with careful scholarship. Precisely because Nicolì's work maps this out clearly, it is an ideal testing-ground for a model-assisted extension – hypothetical, in a future joint effort – without displacing human judgment.

In that workflow, cross-lingual embeddings would search for the corridors Nicolì identifies – across books, journals and anthologies – where wording diverges but meaning holds. The model would propose candidates; the historian would check page images and paratexts, then classify the link (faithful translation, selective appropriation, hostile inversion, pedagogical paraphrase, and also inadvertent, over-hasty misreadings). Where Nicolì documents a path, retrieval can widen coverage to parallel renderings elsewhere; where she notes a singleton, semantic matching can test whether it truly stands alone. The aim is not automation but reach and speed: enlarge the search space and modify the interpretation.

A strong case for this approach is that of the two suppressed essays – 'Of suicide' and 'Of the immortality of the soul' – first printed in French (1770) within the circle of Paul Thiry, baron d'Holbach. Traditional methods get us started; models can add breadth, surfacing reprints, retranslations, and bowdlerized excerpts that would take weeks to find manually. Today, the machine cannot reliably discern consent, intention, or tone; it can, however, flag plausible sites of engagement. Our agenda is to go further – developing cautious proxies (paratextual moves, stylistic shifts, systematic substitutions) so that intention becomes a testable, evidence-based hypothesis. Finally, where compilers domesticated Hume for distinct audiences – stripping out heresy, duplicating pieces in rival versions, omitting allegories – meaning-level matching can group variants, lexical audits can measure moralizing substitutions, and page-level links can show how arguments were redirected. Outputs remain hypotheses, ranked and inspectable, instead of scholarly disagreement, we have a change of thresholds or rules.

The result is that instead of merely tracing instances of Hume's writings across the Channel, we can begin to contextualize him in that setting – by identifying similar cases of domestication by the same

⁶ Transformer-based language models are powerful. When calibrated to 18th-century texts they can also, for example, estimate the publication date (see Rastas, Ryan, Tiihonen, Qaraei, Repo, Babbar, Mäkelä, Tolonen and Ginter 2022).



publisher, for instance – and work with our AI companion to assess how exceptional his case truly was and what that implies for interpretation. Beyond that, we can start to examine the textual environments in which these translations or reused fragments appear – especially when they take the form of short translations or paraphrases rather than full works. By developing systematic methods for analyzing these surrounding contexts, we push interpretation to a new level: not just identifying where and when a passage was used, but understanding how it functioned in its local setting – framed by editorial commentary, embedded in theological or philosophical argument, or aligned with other cultural agendas. These contextual signals help to reveal how ideas were adapted, appropriated, or resisted in ways we have not previously been able to track at scale. This is precisely the kind of interpretive work that would not be possible without computational hermeneutics.

5. Ask ChatGPT (and then what?)

It is increasingly clear that LLMs substantially change the kinds of questions historical research can meaningfully pursue. This is not about querying a chatbot – it is about modeling how meaning evolves across languages, genres, and centuries. Concepts like *sovereignty*, *luxury*, or *nation* can now be traced, not through isolated keywords, but as evolving constellations of paraphrase, citation, and cultural adaptation. Our methods also reveal how Enlightenment ideas appear in places that are easy to overlook: a local election tract quoting Montesquieu, or a sermon reworking Franklin’s economic thought. Of course, it is possible to follow a workflow pattern in which the model proposes and the historian interprets. But the real transformation lies in the ability to move across different resolutions – from individual expressions to system-level dynamics – and to return with sharper insights. This is where, for example, translation mining becomes more than a super-charged finding aid. It reframes the level at which we pose questions. At small scale, we can tell persuasive narratives of reception; at medium scale, we can trace the formation of canons. But only at macro scale do we begin to see who someone like Hume was in the 18th-century information economy and how he later became a figure repeatedly revisited and challenged: which of his arguments were systematically translated, which were watered down by paraphrase or distorted by editorial framing, and which never travelled at all. Alignment models allow us to ask: How did Hume’s political economy circulate relative to his moral essays? Which genres carried him? Which publics engaged with him – and how? The result is comparative intellectual history that begins at scale – across languages, genres, and media – and only then descends to the page level to refine or overturn what we thought we knew.

Thus, perhaps paradoxically, even as the models become larger and technically more powerful, the central challenge shifts from computation to scholarly imagination – toward formulating questions that can fully engage with what these methods now make possible. Thinking about genre is central. Hume himself wrote as both philosopher-essayist and historian – and his reception often split along that seam. *Essays*



and *History* do not simply convey the same claims in different formats; they invite different modes of use, translation, and controversy. A genre-aware pipeline makes that visible. We can train a classifier on Hume's own genre palette and track where 'essay-like' segments cluster within the *History*. My motive for such work is the hope that soon we begin to learn (beyond Hume) why certain lines of philosophical argument traveled in newspapers, histories, and sermons while others remained lodged in collected works (Buntiakova 2025). That is not a pedantic distinction; it is the difference between philosophical argument as public pedagogy (the essay) and as narrative authority (the history). Once measured, the contrast becomes analytically revealing: the same authorial voice moving through channels, relays, and forms of recoding that are far from straightforward.

The 18th-century luxury debate offers another testing-ground for this approach. In our research group, we have sought to lower the threshold between teaching and research by involving advanced students in research on luxury, demonstrating how the approach operates in practice (Hovland 2025). Start with the largest feasible slice of the 18th-century record (e.g., all passages mentioning 'luxury' and its near synonyms across corpora). Instead of leaping straight into opaque modeling, we stage interpretation through definition generation. LLMs are fine-tuned to rewrite each passage as a short, period-sensitive definition of 'luxury' in context; those definitions are then embedded and clustered. Because the clusters live in a transparent 'definition space' – built from the weighted keywords that define them – we can actually read what each cluster represents: civic humanist critiques; theological condemnations; politeness-commerce defenses; fiscal-military arguments; nationalist polemics; moral-sense reframings. The method is fairly robust in the face of OCR messiness and register shifts, but, crucially, it remains inspectable: the clusters are tied to particular passages and their trajectories can be plotted across decades, genres, and languages.

This is also how using Hume in the classroom can meet research without compromising either. Masters students from digital humanities and language technology work on constrained, interpretable data – genre classification within Hume's corpus; detection of essay-like bursts inside the *History*; definition-based clustering of 'luxury' – and that work links to our research group's objectives. The students learn to articulate historical questions as more formal tasks; the group gains validated components that scale. That is not just accelerated scholarship; it is a different scholarship – one that lets us begin at the map, then drill down to the itinerary. And because the whole apparatus is genre-sensitive, it clarifies the boundary between 'philosophy in essays' and 'philosophy in history'. We can hope to show, for instance, where Hume's essayistic rhetoric about refinement and commerce is echoed, softened, or sharpened inside the *History*; where it is stripped of polemic and repackaged as narrative explanation; and how those two registers perform differently once translated or excerpted.

Newspapers and periodicals supply a complementary vantage point on public discourse. For us, early mapping of language, location, and form in Finnish newspapers (1771–1917) offered a template



for delineating national public spheres; later work introduced embedding-based tracking to follow the semantics of core political terms across multiple languages and long spans of time (Tolonen et al. 2019; Hengchen et al. 2021). The methodological lesson carries straight through to Enlightenment studies: newspaper collections – noisy, uneven, but indispensable – can be treated as partners for curated book corpora; we can analyze how reviews, advertisements, extracts, and polemics mediate the birth, life, and afterlives of books; and, crucially, make genre and register shifts within composite texts analytically legible. The humanist questions can be classical – what did ‘nation’, ‘luxury’, or ‘commerce’ come to mean? – but the evidentiary footing is new, because trajectories can be confronted with counter-examples, aligned across languages and reproduced. For Hume, this perspective is crucial. We know very little of his early intellectual development. Yet, we know that he wrote essays at the turn of the 1740s meant for publication and it would be sensational to discover that some of those were actually printed. But systematically mapping his essays onto newspapers lets us move beyond treasure-hunting for the unknown pre-1742 pieces (valuable as that is) to chart the broader ecology of dissemination: reprints and paraphrases of later essays circulating in provincial papers long after he dropped them from his collected works; the *History of England* appearing as a serial publication for readers who could not perhaps afford the books; and, in the background, the *Spectator*-to-miscellany tradition in which ‘essays in newspapers’ and ‘essays in volumes’ continually feed one another. Put together, a scalable newspapers-to-books workflow makes Hume’s public presence measurable – where, when, and in what genres he was read – and restores periodical culture to the center of 18th-century intellectual history rather than treating it as a footnote to the canon. I can already imagine taking the same approach to 19th-century materials to examine the reception of canonical Enlightenment figures, but that work still lies ahead.

A further strand concentrates on historical language and stylistic change. Using corpora like ECCO, we have measured how registers and genres evolved, how economic vocabulary entered common discourse, and how stylistic shifts unfold within individual works (Zhang et al. 2022). Here, Hume’s *Political Discourses* becomes a useful test case: not as ‘the epitome’ of 18th-century economic debate, but as a highly visible node within a much denser argumentative web. The question is not whether the *Political Discourses* stand alone at the summit, but how their phrasing, examples, and argumentative turns relate to adjacent genres – pamphlets, sermons, periodical essays – and to contemporaries who pushed similar claims in different idioms (Tiihonen et al. 2022).

6. Future commitment

Beneath our efforts to understand Hume lies a broader infrastructural commitment: to transform ad hoc computation into sustainable, transparent scholarly practice. Wherever licensing permits, we keep workflows open – documenting how we group editions using metadata, align texts across languages, detect reuse, cluster ornaments, estimate dates, and segment layouts. But openness on its own is not enough.



We build systems that interoperate: national bibliographies remain essential starting points, but we correct their inward-facing biases by harmonizing them and linking them to full text and page images.⁷

What, concretely, does this approach change for Enlightenment studies? It relocates explanation. Canon and corpus need each other. Canonical figures provide explanatory anchors for understanding prestige, authority, and institutional reproduction; corpus control and methods prevent anecdote and reveal long tails – unexpected appropriations, structural biases, and the mundane constraints that shape circulation. Publishers become explanatory actors rather than background; prices are treated as causes, not symptoms; newspapers become mediators rather than mere scenery; and translation is no longer assumed to be faithful or linear but is shown to be selective, hybrid, and often strategically distorted. In practical terms, it also changes pedagogy. The aim is not to produce graduates who can ‘run a topic model’, but to train scholars who understand what it means to design a reproducible study that proceeds from corpus control to uncertainty-aware inference, and from there to defensible interpretation. In that classroom, canonical authors are not mascots for a new method; their works and contexts function as historical hypotheses to be tested across corpora and across languages.

The next step is neither another algorithmic novelty nor a nostalgic return to small-N reading. We need three things: shared benchmarks, modular infrastructure, and model-in-the-loop interpretation that treats the machine as a fallible co-worker with different abilities. Concretely, benchmarks should be based on real historical questions – aligned corpora across languages, gold-standard annotation where possible, and tasks that go beyond generic search. There is still much to do in developing AI models that assist with the contextual understanding of engagement – for instance, in reception studies – but this is just one part of a broader agenda. In parallel, we can envision a stable environment for Enlightenment studies in which books, pamphlets, journals, newspapers, and other digitized materials are regularly added, workflows rerun, and results are compared so that historical claims remain traceable and reproducible. It is not a panopticon, but an ecosystem designed to evolve alongside its community. In such a setting, books are treated at once as material artefacts (images, ornaments, formats), as data objects (structured and unstructured fields), and as vehicles of meaning (arguments, genres, translations). The triadic model – objects, data, meaning – thus becomes operational.

This future is continuous with Enlightenment studies rather than a repudiation. Canonical authors remain central, not as idols, but as nodes in a network the structure of which can be described. The difference is that we now possess instruments that let us scale our questions credibly. Rather than relying on a handful of texts and citations, we can trace continuity and change across hundreds of thousands of works, multiple languages, and centuries. Instead of anecdotal claims about reception, we map structured

⁷ Since the beginning of the COMHIS initiative, one vision of bibliographic data has been to use the *Heritage of the Printed Book* database at CERL as a foundation for building a system that integrates the print culture of early modern Europe into a dynamic, expanding infrastructure, ready to scale globally as new data becomes available.



relationships between ideas, translations, and reuses. We are able to show the page where a connection occurs. Instead of treating books solely as texts, we analyze them as dynamic data objects whose production histories, circulation networks, and semantic afterlives can be modeled. The result is not an abdication of interpretation but an amplification of it. The machine proposes proximities; the scholar judges relevance and then they move forward together. The loop between the two is where explanation and understanding now resides.

The relevance of this transformation extends beyond the academy. In an age shaped by AI, misinformation, and engineered attention, the capacity to understand how ideas change – and how they are reframed or misused – has civic weight and a serious potential for economic uptake. What matters is not only what was said, but how it was interpreted, repackaged, and redeployed. These are the basic interests of any humanist. Rigorous, interpretable AI gives intellectual history a public function: it equips us to identify the mechanics of recontextualization rather than argue endlessly about print editions. That is not a call to automate judgment; it is a demand to make judgment accountable, reproducible, and legible.

This also requires that we learn from the cautionary tale of modern platforms. For a decade, much of digital humanities and computational social science treated Twitter (and later Instagram and other social media platforms) as if they were stable commons. They were not. Access regimes changed; APIs closed; terms of use shifted overnight. What appeared to be ‘open’ turned out to be gatekept by revocable permissions governed by corporate priorities, and entire research programs lost their evidentiary footing. These are what I would call closed continuities: streams that feel continuous until the valve shuts. We should not build our epistemology on sand.

By contrast, historical corpora – however messy – are potentially more durable. Their legal status can be clarified; their provenance documented; their versions archived and cited. Even paywalled collections such as ECCO can be wrapped in workflows that record coverage, gaps, and corrections, so that claims are traceable and disagreements make sense. That durability matters for theory: it lets us ask large, longitudinal questions about public discourse – how polemics travel, how genres domesticate arguments, how crises reorder vocabularies – without the trap of platform volatility. It also matters for ethics: 18th-century actors are dead; the political stakes have cooled with time; the risks to living subjects are minimal. Historical data is not a consolation prize for those denied today’s feeds; it is the best testing-ground we have for understanding communication at scale.

And there is a broader lesson here. We will not predict the past, and we should not pretend to predict the future. But we can model processes – diffusion, translation, appropriation – and study their constraints under conditions we can actually control. Doing so sharpens our ability to read the present without being held hostage to the next API change. All data is historical data; the only differences are latency and custody. If we take that seriously, computational hermeneutics becomes more than a method. It becomes a discipline for thinking with evidence in a world where streams can close tomorrow.



None of this diminishes the craft of intellectual history as we ought to understand it. On the contrary, it puts that craft to more demanding use. Close reading does not disappear; it acquires a new meaning. Archival work is not replaced; it is prioritized where models reveal lacunae that matter. The historian's judgment becomes more – not less – central, because we must decide which model-proposed proximities to ratify as meaningful. We have, in other words, returned to first questions with better instruments: What is an idea? How do ideas evolve? How do they travel, translate, and transform? And how can we model these processes across languages, formats, and centuries without losing the depth that makes humanistic inquiry worth pursuing?

If the field embraces this agenda of computational hermeneutics – benchmarks set in line with real research questions, infrastructures that reconcile national datasets with transnational corpora, and modeling workflows that keep interpretation inside the loop – computational history will cease to be a technical annex and become integral to intellectual history itself. It is the method by which we can honor the Enlightenment's ambitions while acknowledging its limits. And if that sounds ambitious, it should. The alternative is a world of anecdotes, uninspected assumptions, and ever narrower seminar-room debates – contexts in which Hume (for example) can be central only because we already assume he is. The value of computational hermeneutics is that it lets us discover why – and when – such canonical figures endure, how they are made and unmade, and where their shadows fall.

References

- Abhishek D., Bergel G. and Zisserman A.** 2021. 'Visual analysis of chapbooks printed in Scotland'. In: *HIP '21: The 6th International Workshop on Historical Document Imaging and Processing*. New York: ACM Digital Library. <https://doi.org/10.1145/3476887.3476893>.
- Barthes R.** 1977. 'The death of the author'. In Heath S. (trans. and ed.) *Image, Music, Text*. London: Fontana Press, 142–48.
- Buntiakova V.** 2025. 'Exploration of genres in David Hume's *History of England* using machine learning approaches', MA thesis, University of Helsinki.
- Emerson R.** 2008. *Academic Patronage in the Scottish Enlightenment: Glasgow, Edinburgh and St Andrews Universities*. Edinburgh: Edinburgh University Press.
- Fantoli M., Suomela J., van Hal T., Depauw M., Virkki L. and Tolonen M.** 2025. 'Quantifying the presence of Ancient Greek and Latin classics in early modern Britain'. *Journal of Cultural Analytics* 10:1. <https://doi.org/10.22148/001c.128008>.
- Garrett A. and Sebastiani S.** 2017. 'David Hume on race'. In Zack N. (ed.) *The Oxford Handbook of Philosophy and Race*. Oxford: Oxford University Press, 31–43. <https://doi.org/10.1093/oxfordhb/9780190236953.013.43>.
- Garrett D.** 2025. 'Fifty years of scholarship in *Hume Studies*'. *Hume Studies* 50, 13–39. <https://dx.doi.org/10.1353/hms.2025.a958190>.
- Hanisch T.** 2007. 'Using relevance and reception within a contextualist approach'. *Redescriptions: Political Thought, Conceptual History and Feminist Theory* 11, 146–77. <https://journal-redescriptions.org/articles/10.7227/R.11.1.9>.
- Harris J.** 2015. *Hume: An Intellectual Biography*. Cambridge: Cambridge University Press.



- Hengchen S., Ros R., Marjanen J. and Tolonen M.** 2021. 'A data-driven approach to studying changing vocabularies in historical newspaper collections.' *Digital Scholarship in the Humanities* 36: Suppl.2, ii109–ii126. <https://doi.org/10.1093/llc/fqab032>.
- Hinderks K., Ledins C., Ginter F. and Tolonen M.** Forthcoming. 'Translation mining: an AI-driven taxonomy of eighteenth-century Anglo-French translation practices.' *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. <https://doi.org/10.1080/01615440.2026.2675558>.
- Hont I.** 2005. *Jealousy of Trade. International Competition and the Nation-State in Historical Perspective*. Princeton: Princeton University Press.
- Hovland V.** 2025. 'Large Language Models for historical meaning: lexical diachronic semantic change detection in the eighteenth-century luxury debate', MA dissertation, University of Helsinki.
- Ijaz A. Z., Tolonen M., Lahti L. and Tiihonen I.** 2019. 'Analytical determination of editions from bibliographic metadata'. In: Jantunen J. H., Brunni S., Kunnas N., Palviainen S. and Västi K. (eds) *Proceedings of the Research Data and Humanities (RDHUM) 2019 Conference: Data, Methods and Tools*. *Studia Humaniora Ouluensia* 17, 9–19. <http://jultika.oulu.fi/Record/isbn978-952-62-2321-6>.
- Jauss H. R.** 1974. 'Literary history as a challenge to literary theory'. In: Cohen R. (ed.) *New Directions in Literary History*. Baltimore: Johns Hopkins University Press, 11–42.
- Kaukua J. and Lähteenmäki V.** 2020. 'On the standards of conceptual change'. *Journal of the Philosophy of History* 14, 183–204. <https://doi.org/10.1163/18722636-12341418>.
- Kidd C.** 2025. 'The Old World and the intellectual history of American slavery'. *History of European Ideas*, 1–11. <https://doi.org/10.1080/01916599.2025.2597094>.
- Koikkalainen P.** 2011. 'Contextualist dilemmas: methodology of the history of political theory in two stages'. *History of European Ideas* 37, 315–24. <https://doi.org/10.1016/j.histeuroideas.2010.10.010>.
- Kuukkanen J.-M.** 2008. 'Making sense of conceptual change'. *History and Theory* 47, 351–72.
- LaCapra D.** 1980. 'Rethinking intellectual history and reading texts'. *History and Theory* 19, 245–76.
- Lahti L., Marjanen J., Roivainen H. and Tolonen M.** 2019. 'Bibliographic data science and the history of the book (c. 1500–1800)'. *Cataloging & Classification Quarterly* 57:1, 57–78. <https://doi.org/10.1080/01639374.2018.1543747>.
- Mahadevan A., Mathioudakis M., Mäkelä E. and Tolonen M.** 2025. 'Text reuse in large historical corpora: insights from the optimization of a data science system'. *International Journal of Data Science and Analytics* 20:5, 4631–43. <https://doi.org/10.1007/s41060-025-00742-x>.
- Mäkelä E., Singh D., Misson J. and Tolonen M.** 2025. 'Opening the black box of EEBO'. *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/llc/fqaf086>.
- Maslen K.** 1973. *The Bowyer Ornament Stock*. Oxford: Oxford Bibliographical Society.
- Melve L.** 2006. 'Intentions, concepts and reception: an attempt to come to terms with the materialistic and diachronic aspects of the history of ideas'. *History of Political Thought* 27:3, 377–406.
- Nicoli L.** 2025. "'Aussi hardi . . . qu'aucun philosophe en France": the eighteenth-century French reception of Hume's *Essays*'. In: Skjönsberg M., Waldmann F. (eds) *Hume's Essays: A Critical Guide*. Cambridge: Cambridge University Press, 52–70.
- Pocock J. G. A.** 1987. 'The concept of a language and the *métier d'historien*: some considerations on practice'. In: Pagden A. (ed.) *Ideas in Context: The Languages of Political Theory in Early-Modern Europe*. Cambridge: Cambridge University Press, 19–38.



- Prescott A.** 2023. 'Mixed methods and the digital humanities.' In: Schneider B., Löffler B., Mager T. and Hein C. (eds) *Mixing Methods: Practical Insights from the Humanities in the Digital Age*. Bielefeld: Bielefeld University Press, 27–42.
- Rasmussen D.** 2005. *The Problems and Promise of Commercial Society: Adam Smith's Response to Rousseau*. University Park: Penn State University Press.
- Rastas I., Ryan Y. C., Tiihonen I. L. I., Qaraei M., Repo L., Babbar R., Mäkelä E., Tolonen M. and Ginter F.** 2022. 'Explainable publication year prediction of eighteenth century texts with the BERT model.' In Tahmasebi N., Montariol S., Kutuzov A., Hengchen S., Dubossarsky H. and Borin L. (eds) *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*. Dublin: Association for Computational Linguistics, 68–77. <https://aclanthology.org/2022.lchange-1.7.pdf>.
- Ricoeur P.** 1971. 'The model of the text: meaningful action considered as a text', *Social Research* 38, 529–62.
- Robertson J.** 2009. *The Case for The Enlightenment: Scotland and Naples 1680–1760*. Cambridge: Cambridge University Press.
- Rosson D. E., Mäkelä E., Vaara V., Mahadevan A., Ryan Y. C. and Tolonen M.** 2023. 'Reception Reader: exploring text reuse in early modern British publications.' *Journal of Open Humanities Data* 9. <https://doi.org/10.5334/johd.101>.
- Ryan Y. C. and Tolonen M.** 2024. 'The evolution of Scottish Enlightenment publishing.' *Historical Journal* 67:2, 223–55. <https://doi.org/10.1017/S0018246X23000614>.
- Sher R.** 2006. *The Enlightenment and the Book. Scottish Authors and their Publishers in Eighteenth-Century Britain, Ireland, and America*. Chicago: Chicago University Press.
- Skinner Q.** 1966. 'The limits of historical explanations.' *Philosophy* 41, 199–215.
- Skinner Q.** 1969. 'Meaning and understanding in the history of ideas.' *History and Theory* 8, 3–53. <https://doi.org/10.2307/2504188>.
- Spencer M.** 2013. *David Hume: Historical Thinker, Historical Writer*. University Park: Pennsylvania State University Press.
- Spencer M. and Tolonen M.** 2025. 'The reception of David Hume's *Essays* in eighteenth-century Britain.' In: Skjönsberg M. and Waldmann F. (eds) *Hume's Essays: A Critical Guide*. Cambridge: Cambridge University Press, 15–35. <https://doi.org/10.1017/9781009047227.003>.
- Thompson M. P.** 1993. 'Reception theory and the interpretation of historical meaning.' *History and Theory* 32:3. <https://doi.org/10.2307/2505525>.
- Tiihonen I., Lahti L. and Tolonen M.** 2024. 'Print culture and economic constraints: a quantitative analysis of book prices in eighteenth-century Britain.' *Explorations in Economic History* 94. <https://doi.org/10.1016/j.eeh.2024.101614>.
- Tiihonen I., Liimatta A., Pivovarova L., Säily T. and Tolonen M.** 2023. 'Measuring the distribution of Hume's Scotticisms in the ECCO collection.' In: *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*. Dublin: Association for Computational Linguistics, 36–44.
- Tiihonen I., Ryan Y., Pivovarova L., Liimatta A., Säily T. and Tolonen M.** 2022. 'Distinguishing discourses: a data-driven analysis of works and publishing networks of the Scottish Enlightenment.' In: Berglund K., La Mela M. and Zwart I. (eds) *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Conference (DHNB 2022)*. <https://ceur-ws.org/Vol-3232/paper09.pdf>.
- Tiihonen I. and Tolonen M.** 2026. 'Saturated by commerce: a computational analysis of eighteenth-century British political discourse.' *History of European Ideas*, 1–30. <https://doi.org/10.1080/01916599.2025.2609084>.



- Tolonen M.** 2010. 'Self-love and self-liking in the moral and political philosophy of Bernard Mandeville and David Hume', PhD dissertation, University of Helsinki. <https://doi.org/10.5281/zenodo.31942>.
- Tolonen M.** 2026. 'From text to meaning: historical humanities in the age of AI and interdisciplinarity'. In: d'Hoine P., Kohler D. and Decock W. (eds) *Charting the Future of Historical Humanities*. Lectio 16. Turnhout: Brepols, 139–62.
- Tolonen M., Hill M. J., Ijaz A. Z., Vaara V. and Lahti L.** 2021. 'Examining the early modern canon: *The English Short Title Catalogue* and large-scale patterns of cultural production'. In: Baird I. (ed.) *Data Visualization in Enlightenment Literature and Culture*. Cham: Springer International Publishing, 63–119. https://doi.org/10.1007/978-3-030-54913-8_3.
- Tolonen M., Liimatta A., Pivovarova L., Tiihonen I. and Säily T.** Forthcoming 2026. 'Hume's list of Scotticisms in eighteenth-century British context'. *Hume Studies* 51.
- Tolonen M., Mäkelä E. and Lahti L.** 2022. 'The anatomy of Eighteenth Century Collections Online (ECCO)'. *Eighteenth-Century Studies* 56:1, 95–123. <http://doi.org/10.1353/ecs.2022.0060>.
- Tolonen M., Marjanen J., Hill M. and Lahti L.** 2025. 'Book formats, printing practices and reading habits in early modern Europe'. In: Ames S., Terras M. and Gooding P. (eds) *Library Catalogues as Data: Research, Practice, and Usage*. Holborn: Facet, 101–20.
- Tolonen M., Marjanen J., Vaara V., Kanner A., Mäkelä E., Roivainen H. and Lahti L.** 2019. 'A national public sphere? Analyzing the language, location, and form of newspapers in Finland, 1771–1917'. *Journal of European Periodical Studies* 4:1, 54–77. <https://doi.org/10.21825/jeps.v4i1.10483>.
- Tolonen M. and Ryan Y.** 2026. 'Computational methods in intellectual history'. In: Blau A. (ed.) *Meaning and Understanding in the History of Ideas and Beyond*. Proceedings of the British Academy. Liverpool: Liverpool University Press.
- Vaara V. and Tolonen M.** 2025. 'A quantitative and comparative approach to Royalist and Whig sources in Hume's *History of England*'. *Hume Studies* 50, 345–76. <https://doi.org/10.1353/hms.2025.a976682>.
- Vesalainen A., Tolonen M. and Ruotsalainen L.** 2024. 'Document layout error rate (DLER) metric to evaluate image segmentation methods'. *Machine Learning with Applications* 18, 1–28 (article 100606). <https://doi.org/10.1016/j.mlwa.2024.100606>.
- Wang R., Pivovarova L., Ryan Y. and Tolonen M.** Forthcoming. 'Image reuse in eighteenth-century book history: large-scale data-driven study of headpiece ornament variants'. *Digital Humanities Quarterly*.
- Wilkinson H., Briggs J. and Gorissen D.** 2021. 'Computer vision and the creation of a database of printers' ornaments'. *Digital Humanities Quarterly* 15:1. <http://digitalhumanities.org/dhq/vol/15/1/000537/000537.html#>.
- Wu Y., Shu K., Fischer J. P., Pivovarova L., Rosson D., Mäkelä E. and Tolonen M.** 2006. 'Detecting Latin in historical books with Large Language Models: a multimodal benchmark'. In: *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*. <https://aclanthology.org/2026.eacl-long.245/>.
- Zhang J., Ryan Y. C., Rastas I., Ginter F., Tolonen M. and Babbar R.** 2022. 'Detecting sequential genre change in eighteenth-century texts'. In: Karsdorp F., Lassche A. and Nielbo K. (eds) *Proceedings of the Computational Humanities Research Conference 2022*. CEUR Workshop Proceedings 3290, 243–55. <https://www.scopus.com/pages/publications/85143773189>.