Article



Culture machine: How MetaCLIP codifies culture

new media & society I–18 © The Author(s) 2025

Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/14614448251336429 journals.sagepub.com/home/nms



Luke Munn

The University of Queensland, Australia

Adarsh Badri

The University of Queensland, Australia

Abstract

How is the cultural made computational? CLIP models are a recent artificial intelligence (AI) innovation which train on massive amounts of Internet data in order to align language and image, deploying this 'grasp' of cultural concepts to understand prompts, classify images and carry out tasks. To critically investigate this cultural codification, we explore MetaCLIP, a recent variation developed by Meta. We analyse the model's metadata, a single file of 500,000 terms that aims to achieve a 'balanced distribution' or sufficiently broad understanding of concepts. We show how this model assembles histories, languages, ideologies and media artefacts into a kind of cultural knowledge. We argue this codification fuses the ancient technique of the *list* with a more recent technique of *latent space*. We conclude by framing these technologies as *cultural machines* that exert power in defining and operationalising a particular understanding of 'culture' invisibly and at scale.

Keywords

Artificial intelligence, critical AI, cultural commodification, knowledge infrastructure, STS

Corresponding author:

Luke Munn, Centre for Digital Cultures & Societies, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia. Email: l.munn@uq.edu.au

Introduction

When humans are confronted with an image or hear a phrase, they draw on their cultural understanding to make sense of it. This understanding is immensely varied: the morass of people, places, events, ideologies and practices that constitute 'culture' is both wide-ranging and highly disparate. This makes culture notoriously fuzzy or nebulous, a concept Williams (1976) described as one or two of the most complex. And yet codifying the cultural into an 'understanding' is needed if computational models are to adequately grasp human-made cultural artefacts and respond in appropriate ways.

This article thus asks how the cultural, seemingly limitless and indefinable, is rendered computational. The technology industry's answer is CLIP, a recent artificial intelligence (AI) innovation that trains on massive amounts of Internet data in order to align language and image, deploying this 'grasp' of cultural concepts to understand prompts, classify images and carry out tasks (Radford et al., 2021). While machine vision has received some scholarly investigation (Dobson, 2023; Parikka, 2023), there is almost (excepting Amoore et al., 2024) no critical humanities work on this new and highly influential generation of zero-shot multimodal models.

These models are powerful but often inscrutable, black boxes rendered opaque through proprietary techniques, technical expertise and unknown training data. To unpack these systems, we focus on a recent high-profile model: MetaCLIP (Xu et al., 2024). Developed by Meta, Facebook's parent company, MetaCLIP seeks to address and improve on the original CLIP model, particularly in regard to a transparent and replicable development model. The key innovation is the model's 'metadata' file, a single file of 500,000 terms that aims to achieve a 'balanced distribution' of data, supporting a known and sufficiently broad understanding of concepts (Xu et al., 2024). This file is publicly accessible, providing a rare access point into this process of cultural codification and operationalization.

Our analysis focuses on process rather than bias: *how* cultural knowledge is constructed computationally. Bias, while certainly important, has become a dominant and *de facto* mode of cultural critique in relation to AI (Karpouzis, 2024; Peters, 2022; Rozado, 2023). Instead, we take inspiration from Benjamin (1935), who recognised that mechanisation was not just a technical but a cultural transformation. He carefully analysed these technologies and their implications, demonstrating how the new ability to copy and circulate images disconnected them from their original context and the very notion of 'originality'. As he and others (Crary, 1990) have argued, each new wave of technical innovation needs to be accompanied by a societal adjustment that critically grasps how these technologies reshape everyday life.

Our key contribution is to demonstrate how knowledge is codified, formalised and operationalised. While content is certainly key to this process, we aim to identify a general procedure or overall approach to culture, a logic. We show how this model blends the *list* – a highly effective ancient knowledge technology – with *latent space*, a far newer and flexible technology. These two technologies augment each other, contributing to a powerful (but highly particular) kind of 'cultural understanding'.

These models matter. As daily life becomes increasingly digitally mediated, the form of 'culture' registered and reproduced by technologies also grows in importance.

Everyday cultural practices on platforms are shaped by the models and algorithms that underpin them (Bucher, 2018; Striphas, 2015). For instance, AI recommendation systems not only predict but influence people's preferences and choices (Erdoğan, 2023; Gaw, 2022), shaping the distinct way in which they see, understand and know culture (Seaver, 2019, 2022). These technical processes construct a kind of cultural feedback loop, with users, communities and tastemakers adjusting their practices to maximise legibility and exposure (Gillespie, 2014; Munn, 2018). In this sense, (as our conclusion elaborates) models like MetaCLIP function as *cultural machines*, defining and operationalising a certain understanding of 'culture' in real time and at scale. Such machines exert political power, elevating some people, practices, ideas and events while marginalising others.

Literature review: from technical systems to cultural machines

Models like CLIP are multifaceted objects: technical systems, composed of algorithms and shaped by training data, that codify knowledge and exert cultural force in the world. This section thus draws together disciplinary perspectives – infrastructure studies, critical AI and algorithm studies, and cultural studies – to frame this object and situate our intervention in relation to it.

As technical systems, AI models are often framed as black boxes, opaque or inaccessible objects that are intimidating to non-specialists. This opacity can be deliberate, trade secrets that shroud products from the public and make deconstruction difficult (Pasquale, 2015). Software studies (Fuller, 2008; Marino, 2009) have suggested source code readings and reverse engineering could advance our understanding of how technical systems operate. Similarly, infrastructure studies (Munn, 2023; Parks and Starosielski, 2017) and media archaeology (Kirschenbaum, 2008) have attempted to make the invisible visible, stressing these ubiquitous and banal systems actually shape our everyday life in powerful ways.

Responding to this opacity, researchers have proposed interventions to cultivate transparency. Model cards provide insights into a model's training data and social, cultural and intersectional context (Mitchell et al., 2019). Explainable AI seeks to provide clarity to users about how these systems make decisions (Holzinger et al., 2022; Minh et al., 2022). Frameworks and toolkits to better understand AI technologies – particularly their democratic potential (Swist et al., 2023) and ethical usage (Adams and Groten, 2023) – have been advanced. While not our primary focus, our work mirrors such studies in using publicly accessible information to unpack the data and decision-making inside these often inscrutable objects.

A dominant strand of AI scholarship has concentrated on AI bias or algorithmic bias. Training data, code rules and statistical infections can all create forms of prejudice in cultural, political or racial forms (Devillers et al., 2021; Fazil et al., 2023; Karpouzis, 2024; Leavy et al., 2020; Manasi et al., 2022; Peters, 2022; Rozado, 2023). While biases certainly matter, the de facto move of analysing and identifying bias can short-circuit the more extended project of understanding a model's logic and its implications.

In this article, we focus on these larger questions: what does 'culture' look like from a machinic perspective, how does a model approach and internalise this knowledge and what are the implications for deploying these cultural machines in the wild? Rather than a top-down approach – measuring a model against some normative yardstick of culture – we take a bottom-up approach that aims to grasp how 'culture' is computationally understood and operationalised. In this sense, we are interested not in predefined notions of culture (which has been extensively theorised), but in how influential new technologies like AI models approach it and render it digital and performable.

Culture is notoriously nebulous, one 'of the two or three most complicated words in the English language' (Williams, 1976: 76) and sprawls across 'every society and every mind' (Williams, 2011: 55). Like Foucault (2005: 172), we are fascinated by how one might take a 'whole domain' of culture and render it 'describable and orderable'. Or, following Goody (1977: 94), we ask how these models resolve this problem in a way that produces 'increments of knowledge' and the 'organisation of experience'.

Certainly others have pursued similar inquiries, most notably in scholarship on algorithmic culture (Dourish, 2016; Hallinan and Striphas, 2016; Hristova et al., 2020; Hutchinson, 2017). As Striphas (2015: 395) notes, over the last few decades, 'human beings have been delegating the work of culture – the sorting, classifying, hierarchising of people, places, objects, and ideas increasingly to computational processes'. Seyfert and Roberge (2016: 3) stress that algorithms have 'agency and performativity'; in sorting and classifying this material in particular ways they bring to life certain visions while leaving behind others. In short, algorithms carry out work in the world, and this work has downstream impacts.

Of course, algorithms are not entirely deterministic. As Andersen (2020) reminds us, we treat these technologies like other forms of communication, making them fit into our everyday lives. Algorithmic cultures may be accepted or overwritten, as different definitional frameworks or understandings of culture jockey with each other (Hallinan and Striphas, 2016). Indeed, in the last decade, we have seen how some algorithms become 'culturally meaningful' objects in their own right: "data" to be debated or tracked, legible signifiers of shifting public taste or a culture gone mad, depending on the observer' (Gillespie, 2016: 64).

We build on these critical insights while also striving to tease out the novel aspects of recent AI models like CLIP. Certainly at one level, AI models are inherently algorithmic, with algorithms like backpropagation a key part of machine learning. However, our work with MetaCLIP (discussed below) highlights at least two points of departure that are at once technical and epistemological. First, rather than imposing rules on cultural content, AI models begin from this messy material, aiming to extract patterns and discover unexpected correlations from them. In this sense, they reverse the flow stipulated by algorithmic culture, starting from repertoires of thoughts and structures of feeling (Williams, 1958) and using these to construct computational models.

Second, AI logics move beyond a strictly algorithmic model, conventionally understood as a set of rules or a repeatable recipe (Munn, 2018). In these models, as we later discuss, harder classificatory schemes are augmented by softer 'contrastive' schemes which organise and operationalise information in flexible and open-ended ways. These points suggest that while such AI models undoubtedly build on existing technologies, they cannot be conceptualised merely by relying on earlier work. In this sense, we see AI models (and our work here) as extrapolating from excellent scholarship on algorithmic cultures while attending to some novel dynamics.

Methodology: unmaking Meta's cultural machine

CLIP represents the most recent and significant development in this drive to codify culture. Released in 2021 by OpenAI, CLIP stands for contrastive language-image pretraining. In the words of its developers, CLIP aims to 'test the ability of models to generalise to arbitrary image classification tasks in a zero-shot manner' (OpenAI, 2021). Rather than manually annotating images, which is expensive and time-consuming, or classifying them into a limited set of known categories ('airplane', 'boat', 'person'), CLIP seeks to be a general-purpose model that can automatically grasp new images and dynamically associate them with similar images.

To develop this understanding, CLIP is trained on 400 million image-text pairs scraped from the Internet. In essence, it is trained on 'culture'—a hugely varied set of everyday imagery which captures sports, hobbies, outdoor scenes, tourist landmarks, film stills and a myriad of other objects, scenes and settings. Crucially, this foundational 'knowledge' allows the model to then grasp the contents of new images, data that were never encountered during training. The 'cultural training' embedded in the model allows images to be broken down into components and linked to a textual description that is most similar to it.

However, what exactly is in this training data is unknown. Is it really representative, capturing the full gamut of people, events, objects and ideas that constitute 'culture'? To answer this crucial question, one would need to develop a 'map' of culture with diverse and important landmarks, and then map these 400 million image-texts to these key points. This is effectively what MetaCLIP aims to do. In the words of the developers, it 'takes a raw data pool. . . and yields a balanced subset over the metadata distribution' (Xu et al., 2024). The goal is not only to better grasp the training data, but to ensure that it is spread across the cultural space, providing good-enough 'coverage' of any person, object or event that should be understood. Whatever the cultural element, the model should be able to locate it within its map or distribution; this 'task-agnostic data' is key for any foundation model (Xu et al., 2024: 1).

To explore MetaCLIP, we analyse the metadata that both represents its innovation and gives it its name. This metadata consists of 500,000 terms, separated by commas, contained in a single file. To aid in understanding, we convert the file to a spreadsheet and arrange it alphabetically. Inspired by exploratory research, which focuses on discovery within new domains (Swedberg, 2020), we undertake a series of quick and light analyses of the file, from counting cultural artefacts to determining languages. The aim is not to exhaustively analyse each term (an impossible task given the volume of information), nor to identify bias or limits in the content, but rather to grasp how data are taken up and worked on in certain ways to form a particular kind of 'cultural knowledge'.

Analysis: constructing cultural knowledge

In this section, we survey key elements of MetaCLIP's metadata file, showing how content is brought together to form a particular kind of cultural understanding (Meta, 2024). This file functions as a kind of cultural map (Meyer, 2014), providing insight into what constitutes culture: how societies value different things, what their beliefs and preferences are and how they engage with each other (Anonymous, 2016).

In approaching this file, we employed a version of exploratory data analysis, a more ad hoc approach that aims to get a grasp of the content and scope of a data set (Leinhardt and Wasserman, 1979). Our analysis of scanning, reading and reflecting on the file was deliberately manual rather than automated, an approach which we felt better employed our expertise as media studies and humanities scholars. In particular, we focused on elements that re-occurred in the file with high frequency and could be clustered into overarching categories. From this approach, history, language, media and political ideologies emerged as distinct 'ingredients' or elements. In the following section, we briefly discuss each of these elements and how they contribute to the 'balanced' and comprehensive cultural understanding the model aims for.

Language

To explore what languages are included in MetaCLIP, we carried out a light analysis using the DetectLanguage function on the first 100,000 entries. There are over 110 languages in the metadata file, from Hmong through to Javanese and Bhojpuri. In one sense, then, there is an incredible variety of languages feeding into this cultural knowledge. This resonates with research suggesting loan-words from other languages are one concrete way in which cultural learning takes place (Higa, 1979). Yet, this diversity is rather superficial, with most languages having less than 1000 entries and many only around a dozen. As Figure 1 demonstrates, English dominates all others, with nearly two-thirds of the total entries. The hegemony of English (Macedo et al., 2015) remains unchallenged.

While this dominance certainly matters, the key point here is that language – more basically, semantic constructions of letters and symbols – becomes a critical component in codifying culture. Culture and language do not exist in isolation, but mutually constitute and reinforce each other (Witherspoon, 1980). Language is a key vehicle for cultural semiotics: cultural meaning is encoded in linguistic signs (Kramsch, 2014). Language provides people with a toolkit of symbols and words to interact and engage with culture. In this way, it structures thought and provides a cognitive frame to reflect on the world, enabling us to understand what others value and respect. At a fundamental level, MetaCLIP is predicated on this tight language–culture connection, on the assumption that culture can be understood as a vast array of people, places, events and ideologies, written out and ingested as information.

History

History is essential to grasping culture, providing context to how various practices, beliefs and value systems emerged in a society. History and culture are intertwined



Figure 1. Make up of languages in the first 100,000 entries (only top 50 shown).

(Nunn, 2012). Our understanding of culture follows historical antecedents that logically explain patterns of behaviour and practices within a culture (Histen, 2024). Therefore, history shapes culture and is, in turn, shaped by it. History appears in MetaCLIP's metadata in two key ways (Tables 1, 2). First, the file contains over 300 terms with the word 'history', documenting histories of nation-states (e.g. Australia, Afghanistan, and India), multinational conglomerates (e.g. Apple, Facebook, and Amazon), food and beverages (e.g. pizza, beer, and wine), and technologies (e.g. cryptocurrency, Python, and Microsoft Office). As such, this history tends towards the popular. Sports, cuisines and tech are well documented, whereas the more conventional history taught at school (empires, revolutions, religious practices) barely finds any mention. For instance, the terms 'war' and 'history' appear only once together in the entire file with an entry on the History of the Peloponnesian War.

History also appears in the file in the form of historical events prefixed by year. These 4000 terms begin with the 1900 Galveston Hurricane and end in the present with current events like the 2024 Summer Olympics and future entries such as the 2026 FIFA World Cup. Sports matches and violent events (terrorism, coups, riots) tended to dominate this list. Both categories can be classed as a version of spectacle, a concept Debord (2021) links with the apparatus of media and the stream of images representing daily life. In this sense, this 'history' replicates the viral logic of the Internet, where football goals and violent mobs achieve virality (Wasik, 2009). In both instances – an emphasis on popular rather than 'academic' history and the dominance of sport and violence – we see how MetaCLIP internalises and perpetuates what societies value as history.

Table I. 'History of' entries (excerpt).

History of Facebook History of Faizabad History of Federal Open Market Committee actions History of film History of firefighting History of Florida History of Ford Motor Company History of foreign relations of China History of France History of free and open-source software

 Table 2. Historical events (excerpt).

2002 Bali bombings 2002 FIFA World Cup 2002 FIFA World Cup knockout stage 2002 FIFA World Cup squads 2002 Formula One World Championship 2002 Gujarat riots 2002 Gujarat violence 2002 in film 2002 in music 2002 in video games

While the merits of this limited and spectacle-focused history could certainly be debated, the broader point is that history is a key ingredient in this AI model's construction of cultural knowledge. Since history seeks to maintain a record of culture, the knowledge of culture is (partially) knowledge of key events in the past, including (as here) 'pop events' or 'mundane' events. History does not just compel AI to synthesise what is acceptable knowledge and values in a given society at a specific historical time and space, thereby deeming AI systems to reduce inherent information bias and become sensible machines. Therefore, history as a cultural category helps make sense of how AI understands culture as a whole.

Politics

Politics and culture are intertwined, with culture playing a critical role in shaping politics and political choices (Pye and Verba, 2015). Political cultures encapsulate the set of shared ethical values, beliefs and normative knowledge a society holds about political processes (Berezin, 1997). Similarly, political ideologies provide a set of ideas about how politics should work in societies. All this suggests that any adequate grasp of culture requires a basic understanding of politics or 'isms'.



Figure 2. Word cloud of political ideologies, size indicates frequency (Meta, 2024).

As a basic mapping of political material, we counted which top-level terms from the 'list of political ideologies' (Wikipedia, 2024) appeared in the metadata file (Figure 2). Approximate tallies were democracy (156), colonialism (92), communism (90), environmentalism (89), socialism (75), liberalism (66), feminism (38), anarchism (38), fascism (36), conservatism (36), capitalism (28), nationalism (27), libertarianism (17), syndicalism (8), authoritarianism (7), populism (7), postcolonialism (6), transhumanism (4), and communitarianism (2). The appearance of the terms 'recolonize', 'recolonization' and 'recolonized' was surprising, given the stigma associated with colonisation (Césaire, 2000).

Based on the model's overall logic, these political ideologies become waypoints or hubs, like the other metadata entries, around which the 'raw' data are distributed. In a practical sense, this would allow the model to link the concept of 'democracy' with texts or images of protests or parliamentary proceedings, while 'authoritarianism' could be associated with military parades. In this way, political ideologies provide this AI model with a basic way to map particular politics to various peoples, places, and times, as well as certain aesthetics and symbols. Such interlinked data mirror scholarly insights that culture is inherently political, bound up with institutional power, political orders, and ideological assumptions (Wright, 1998). In this way, politics becomes one of many ingredients that contribute to the 'balanced' or comprehensive cultural knowledge aimed at by the AI model.

Media

Culture is deeply intertwined with expressions of art and entertainment. By displaying music, literature, visual art, films and TV series, societies tend to express their cultural experiences and aspirations. These cultural objects are the products of a cultural industry, a globally distributed sector, producing a vast array of commodities, where old dichotomies of high/low or elite/popular no longer apply (Rodríguez-Ferrándiz, 2014).

MetaCLIP's metadata contains entries on thousands of specific cultural artefacts in arts and entertainment (e.g. *Red Dead Redemption*). Because many are Wiki entries, they list their media type in brackets, such as '(videogame)' or '(TV series)', allowing us to obtain a basic tally of each type. Within the metadata, there are 1585 film entries, 1320 TV series entries, 232 bands, 207 graphic novels, 203 video games and 161 albums. Figure 3 visualises a selection of these artefacts, gesturing to this extensive media landscape.

While we note the Westernised slant to this media, our key point is that these entries suggest artefacts provide a window into a culture. Indeed, the lopsidedness of this material itself reflects a media terrain of American-exports and global consumption practices. Objects embody a 'syntax of cultural practices', presenting 'modes of meaning in the world, dispositions of thought and comportment' in a highly concrete way (Lippit, 2024). MetaCLIP reflects this insight: a key vector to understanding culture is to identify, order and 'decode' many of its key cultural artefacts. At a basic level, speech, humour and texts can often only be understood if one knows the 'cultural reference' it refers to or rifs of. By including a massive array of films, TV series, music albums, novels and video games, the AI model seeks to build cultural understanding via media.

Discussion: from the list to latent space

To codify knowledge, MetaCLIP uses two key techniques: the list and latent space.

The list

MetaCLIP's metadata can firstly be understood as a giant list, an index of 500,000 entries ranging from places to events, ideas and artefacts. Within this list, there are references to other lists. 'List of American Films', for instance, documents American movies according to the year of release, functioning as a digital placeholder for cultural objects. Indeed, the file contains thousands of list entries, from TV series to Academy Awards and NFL Championships. By classifying and developing knowledge domains (Werbin, 2017), these Lists construct cultural knowledge.

The List is both ancient and contemporary. Humans have used lists to document and structure their lives since antiquity, and in this sense, can be viewed as the 'origin of culture' (Eco, 2009). The List has been critical to managing population, territory and security and has served to administer, organise, manage, police and circulate people and things (Werbin, 2017; Young, 2017). Lists are highly flexible knowledge systems with distinct aims: 'past, present or future-oriented; retroactive, administrative, or prescriptive' (Young, 2017: 16). The List has risen to become a key contemporary



Figure 3. Indicative image cloud of media artefacts in the metadata.

technology, with email lists, wish lists, no-fly lists, and kill lists becoming a ubiquitous feature of everyday life (de Goede and Sullivan, 2016; Johns, 2016; Staeheli, 2012; Weber, 2016). Today's lists are increasingly taking the shape of databases and digital files, which have enhanced their ability to structure the world (de Goede et al., 2016). Lists are increasingly 'giving form to protocols and algorithms' that reshape how we think about the world around us (Young, 2017: 14). MetaCLIP as List thus continues its role as a technology for codifying and formalising knowledge.

In the model metadata, the List functions as a 'conceptual prison' (Goody, 1977), a defined space that formalises what culture is. 'List of 2020 albums', for instance, creates a container that is then populated. This container has 'a clear-cut beginning and a precise end, that is, a boundary, an edge, like a piece of cloth' (Goody, 1977:

81). In this sense, lists take something open-ended and give it a form; 'making a "cut" in the continuous flow of the world' (de Goede et al., 2016: 6). 'Lists teach us about systems of order that surround and enframe us', notes Young (2017: 15), providing insights into how information is scaffolded to form systems of knowledge production and circulation. Because culture cannot be captured as a whole in data points, the model uses lists to classify and encode different artefacts into rigid categories. The List produces a meaningful grouping, where each object is placed within a fixed class that is tightly defined.

This boundary-making ability makes the List a 'privileged tool' for rendering culture intelligible (Staeheli, 2012). Once published and circulated, the list disciplines the items and ideas judged list-worthy and those which are not. In the Foucauldian sense, the list generates a 'regime of truth', a claim about things deemed to be a 'fact' and, therefore, irrefutable (Lorenzini, 2015). In this sense, the List no longer remains a tool for classifying and creating new semantic fields (Goody, 1977) but functions as an 'epistemic practice and regulatory technique' (de Goede et al., 2016: 4). In other words, the List is not merely technical but cultural and political. The List is the definitive record of what counts and does not count in culture. In essence, MetaCLIP's use of the List as a technology develops a cultural map, allowing it to make sense of that complex phenomenon we call culture.

Latent space

In MetaCLIP, the List is joined with a far newer technology: *Latent Space*. Latent Space or embedding space simply denotes an abstract, multi-dimensional space that captures essential features of input data. CLIP's key task and its greatest innovation, as discussed earlier, are to align 'language' and 'image' representations in this space. A photograph of a mountain landscape and the phrase 'mountain landscape' are distilled into two numerical values and will lie close to each other in space, indicating their similarity. The utility of latent space is 'at the core of deep learning', writes one computer scientist (Tiu, 2020), 'learning the features of data and simplifying data representations for the purpose of finding patterns'. Despite this importance, there is almost no (excepting Amoore et al., 2024) scholarship in media or cultural studies on latent space.

In one sense, latent space is about compression. A heavy or dense piece of data, such as a high-resolution image, can be distilled down to an 'essential feature' that occupies a particular point in space. Yet, even this technical efficacy has clear epistemic implications. In drastically downsizing original data to an embedded feature, aspects of the original are inevitably lost, but far more data can be processed and internalised. Latent space attempts not to exhaustively encode a handful of cultural artefacts or concepts, but rather to more crudely capture an immense variety of objects and events, places and people.

This strategy has clear implications, suggesting it is the relationship between items that define and indeed supersede its 'essence'. Concepts like '70s pop' or 'the Cold War' are not captured definitionally but spatially – their essence is not housed in a list of precepts but can be found hovering around a particular point in space. This is a very flexible way to internalise concepts and relationships. There is no hard boundary or border

The list	Latent space
classification	contrast
boundaried, closed	free-floating, open-ended
conceptual prison	conceptual realm
flexed, definitive	flexible, malleable
self-contained	interconnected
certain, absolute	equivocal, fuzzy
rationalist	relationalist

Table 3. Key terms differentiating lists and latent space as knowledge technologies.

between concepts, but instead, more or less distance. An event like the 'Cuban missile crisis' may occupy a cluster with other 'Cold War'-related entries, but also lies within some proximity to 'Cuban jazz'. Indeed, as users of ChatGPT or StableDiffusion know, it is possible to explicitly explore these grey zones or overlapping areas that are the spatial midpoint of two concepts.

This spatial logic is about *contrast* rather than *classification* as computer science typically understands it; it is not about sorting data into predetermined and hard-edged categories (Krishnaiah et al., 2014). Entities do not have some innate essence, nor any kind of bright line test required to belong or not belong. Rather, entities derive their identity through contrast: by being more or less like other things. A 'cat' cluster is in closer proximity to 'dog' than 'people'. This spatial logic is what allows techniques like negative prompting to function. In positively prompting 'beautiful' and negatively prompting 'ugly or deformed' – an ableist but all-too-common strategy – users are explicitly pushing towards some points and away from others.

The List and Latent Space thus operate according to fundamentally different logics, a distinction we have gestured to with the keywords above (Table 3). For example, while the List employs a *rationalist* approach where things exist independently, Latent Space takes a more *relational* approach to knowing, a critique of Enlightenment-based empiricism (Kuhn, 1962) that asserts that individuals and societies are co-constitutive (Donati, 2011; Prandini, 2015). Mirroring these insights, the knowledge codified in Latent Space emerges from relations between things. Close proximity creates a fuzzy coherence. In this sense, the Latent Space emphasises constitution, interaction and connection between things. If the List asserts absolute certainty about cultural categories, Latent Space offers a more open-ended and equivocal way-of-knowing.

Conclusion: the political power of cultural machines

In this article, we have explored how MetaCLIP constructs an 'understanding' of culture. In some respects, this use of classifying, relating and ordering continues long-standing patterns of knowledge construction. However, in other ways – the millions of training data points, aligned to hundreds of thousands of metadata entries, the distinct combination of lists and latent space which impose order while allowing for flexibility and relationality – we see how this understanding is computational and (partially) novel. In

amassing and training on this cultural material and deploying this particular understanding of culture in various applications at scale, we suggest MetaCLIP and models like it function as cultural machines.

Such machines can exert force precisely because culture is fluid, not fixed. As Benjamin (1940) stressed, culture is not defined by rules – an idealised and static entity from the past – but is our ever-changing understanding of humans' place in the world. Culture is not 'an essential set of relations between a people, place and way of life' but a 'conjunctural and pliable articulation' based on our creative form-giving capabilities (Bennett, 2015: 555). Cultures are dynamic and interactive (Chemla and Keller, 2017): they are not simply discovered but are actively constructed.

If culture is up for grabs, computational systems increasingly shape it. Algorithmic regimes, big data and machine learning models internalise and instrumentalise massive amounts of user data or harvest it en-masse from that sprawling cultural artefact known as the Internet. By codifying knowledge at scale, these cultural machines grasp (albeit in a limited way) information about people, places, things and events and some of the relationships between them. Increasingly embedded into many aspects of human existence, these computational cultures operate in a dynamic, agile and interactive way (Giannini and Bowen, 2022).

Driven by data, cultural machines make choices about what counts. 'Reasoning about X requires representations of X', Agre (2003: 287) observed; these representations 'need to encode all the salient details' of culture so that 'computational processes can efficiently recover and manipulate them'. While this statement appears logical, the question of salience is deeply political. What matters, what is significant? MetaCLIP's metadata provides one answer: these are the people, places, artefacts, events and ideologies considered key for constructing a 'sufficient' cultural understanding. Certainly this approach is deeply impoverished: affect, interpersonal relationships, oral tradition – all these aspects of culture are left out of this information-centric approach. Yet, salience consciously sheds such detail to produce something actionable. By amassing, ordering and aligning data, models like MetaCLIP transform cultural artefacts into a kind of cultural knowledge.

In defining culture, cultural machines exert political power, a 'distribution of the sensible' (Ranciere, 2013) that deeply shapes what is visible and what is invisible, what is spoken and what is silenced. These models instantiate a model of the world, a particular way-of-knowing that is inherently ethical and political (Amoore et al., 2024). Such infrastructures enact a kind of politics of parameters (Magee and Rossiter, 2015). Baked into neural networks and carried out automatically, they are decisions that never register as decisions. Images are tagged, associations made and posts categorised silently and invisibly. Indeed, we suggest the political power of cultural machines resides precisely in this combination of the banal and the procedural – backend operations. Drawing on insights from across the humanities and social sciences, our work here presents a first step in understanding the logic and limitations of the cultural machines that increasingly mediate our everyday lives.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Luke Munn (D) https://orcid.org/0000-0002-1018-7433

References

- Adams C and Groten S (2023) A TechnoEthical framework for teachers. *Learning, Media and Technology* 49(4): 701–718.
- Agre P (2003) Writing and representation. In: Mateas M and Sengers P (eds) *Narrative Intelligence*. Amsterdam: John Benjamins, pp. 281–304.
- Amoore L, Campolo A, Jacobsen B, et al. (2024). A world model: On the political logics of generative AI. *Political Geography* 113: 103134.
- Andersen J (2020) Understanding and interpreting algorithms: toward a hermeneutics of algorithms. *Media, Culture & Society* 42(7–8): 1479–1494.
- Anonymous (2016) Culture. In: Introduction to Sociology: Understanding and Changing the Social World. Minneapolis, MN: University of Minnesota.
- Benjamin W (1935) The work of art in the age of mechanical reproduction. In: Arendt H (ed.) *Illuminations*. New York: Schocken Books, pp. 1–26.
- Benjamin W (1940) On the concept of history. Available at: https://www.marxists.org/reference/ archive/benjamin/1940/history.htm (accessed 28 August 2024).
- Bennett T (2015) Cultural studies and the culture concept. Cultural Studies 29(4): 546–568.
- Bucher T (2018) If. . . Then: Algorithmic Power and Politics. Oxford: Oxford University Press.

Césaire A (2000) Discourse on Colonialism. New York: NYU Press.

- Chemla K and Keller EF (2017) *Cultures without Culturalism: The Making of Scientific Knowledge*. Durham: Duke University Press.
- Crary J (1990) Techniques of the Observer. Cambridge, MA: MIT Press.
- de Goede M and Sullivan G (2016) The politics of security lists. *Environment and Planning D:* Society and Space 34(1): 67–88.
- de Goede M, Leander A and Sullivan G (2016) Introduction: the politics of the list. *Environment* and Planning D: Society and Space 34(1): 3–13.
- Debord G (2021) The Society of the Spectacle. Cambridge: Unredacted Word.
- Devillers L, Fogelman-Soulié F and Baeza-Yates R (2021) AI & human values. In: Braunschweig B and Ghallab M (eds) *Reflections on Artificial Intelligence for Humanity*. Cham: Springer, pp. 76–89.
- Dobson JE (2023) *The Birth of Computer Vision*. Minneapolis, MN: University of Minnesota Press.
- Donati P (ed.) (2011) *Relational Sociology: A New Paradigm for the Social Sciences*. London: Routledge.
- Dourish P (2016) Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3(2): 1–11.
- Eco U (2009) Spiegel interview with Umberto Eco: 'we like lists because we don't want to die'. Available at: https://www.spiegel.de/international/zeitgeist/spiegel-interview-with-umbertoeco-we-like-lists-because-we-don-t-want-to-die-a-659577.html (accessed 23 June 2024).
- Erdoğan Z (2023) Netflix's machine learning, personalization, culture effect and its evolution in COVID-19. *Intermedia International E-journal* 10(18): 1–14.

- Fazil AW, Hakimi M and Shahidzay AK (2023) A comprehensive review of bias in AI algorithms. *Nusantara Hasana Journal* 3(8): 1–11.
- Foucault M (2005) The Order of Things. London: Routledge.
- Fuller M (2008) Software Studies: A Lexicon. Cambridge, MA: MIT Press.
- Gaw F (2022) Algorithmic logics and the construction of cultural taste of the Netflix Recommender System. *Media, Culture & Society* 44(4): 706–725.
- Giannini T and Bowen JP (2022) Museums and digital culture: from reality to digitality in the age of COVID-19. *Heritage* 5(1): 192–214.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) Media Technologies Essays on Communication, Materiality, and Society. Cambridge, MA: MIT Press, pp. 167–194.
- Gillespie T (2016) #TrendingisTrending: when algorithms become culture. In: Seyfert R and Roberge J (eds) *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies.* Abingdon: Routledge, pp. 52–75.
- Goody J (1977) The Domestication of the Savage Mind. Cambridge: Cambridge University Press.
- Hallinan B and Striphas T (2016) Recommended for you: the Netflix Prize and the production of algorithmic culture. New Media & Society 18(1): 117–137.
- Higa, M. 1979. Sociolinguistic aspects of word borrowing. In W. Mackey and J. Ornstein (eds) Sociolinguistic Studies in Language Contact. The Hague: Mouton, pp. 277–292.
- Histen MJ (2024) History and cultural evolution: measuring the relationship through the Wikipedia network. *Journal of Institutional Economics* 20(18): 1–20.
- Holzinger A, Saranti A, Molnar C, et al. (2022) Explainable AI methods–a brief overview. In: Holzinger A, Goebel R, Fong R, et al. (eds) xxAI–Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers. Lecture Notes in Computer Science. Cham: Springer, pp. 13–38.
- Hristova S, Slack JD, Hong S, et al. (2020) *Algorithmic Culture: How Big Data and Artificial Intelligence Are Transforming Everyday Life.* Lanham, MD: Lexington Books.
- Hutchinson J (2017) Algorithmic culture and cultural intermediation. In: Hutchinson J (ed.) *Cultural Intermediaries: Audience Participation in Media Organisations*. Cham: Springer, pp. 201–219.
- Johns F (2016) Global governance through the pairing of list and algorithm. *Environment and Planning D: Society and Space* 34(1): 126–149.
- Kirschenbaum M (2008) *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: MIT Press.
- Kramsch C. 2014. Language and culture. AILA Review 27(1): 30-55.
- Krishnaiah V, Narsimha G and Chandra SN (2014) Survey of classification techniques in data mining. *International Journal of Computer Sciences and Engineering* 2(9): 65–74.
- Kuhn TS (1962) The Structure of Scientific Revolutions. Chicago, IL: University of Chicago Press.
- Leavy S, O'Sullivan B and Siapera E (2020) Data, power and bias in artificial intelligence. arXiv:2008.07341. arXiv. Available at: http://arxiv.org/abs/2008.07341 (accessed 23 June 2024).
- Leinhardt S and Wasserman SS (1979) Exploratory data analysis: an introduction to selected methods. *Sociological Methodology* 10: 311–365.
- Lippit Y (2024) Understanding culture through material artifacts. Available at: https://vpal.harvard.edu/understanding-culture-through-material-artifacts (accessed 30 August 2024).
- Lorenzini D (2015) What is a 'regime of truth'? Le foucaldien 1(1): 1-5.
- Macedo D, Dendrinos B, and Gounari P (2015) Hegemony of English. New York: Routledge.

- Magee L and Rossiter N (2015) Service orientations: data, institutions, labor. In: Kaldrack I and Leeker M (ed.) *There Is No Software, there Are just Services*. Leuphana: Meson Press, pp. 73–89.
- Manasi A, Panchanadeswaran S, Sours E, et al. (2022) Mirroring the bias: gender and artificial intelligence. *Gender, Technology and Development* 26(3): 295–305.
- Marino M (2009) Critical code and software studies. *Digital Humanities* 2009: 19–20. Available at: https://eis.ucsc.edu/papers/nwf-CS11-dh09-expressiveProcessing.pdf (accessed 31 August 2024).
- Meta. (2024) *MetaCLIP Metadata*. Available at: https://raw.githubusercontent.com/facebookre-search/MetaCLIP/main/metadata.json
- Meyer E (2014) *The Culture Map: Breaking through the Invisible Boundaries of Global Business*. New York: Public Affairs.
- Minh D, Wang HX, Li YF, et al. (2022) Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review* 55(5): 3503–3568.
- Mitchell M, Wu S, Zaldivar A, et al. (2019) Model cards for model reporting. In: Conference on fairness, accountability, and transparency, Atlanta, Georgia, 29–31 January, pp. 220–229. New York: Association for Computing Machinery.
- Munn L (2018) Ferocious Logics: Unmaking the Algorithm. Lüneburg: Meson Press.
- Munn L (2022) Countering the Cloud: Thinking with and against Data Infrastructure. London: Routledge.
- Munn L (2023) Technical Territories: Data, Subjects, and Spaces in Infrastructural Asia. Ann Arbor, MI: University of Michigan Press.
- Nunn N (2012) Culture and the historical process. *Economic History of Developing Regions* 27(Suppl. 1): S108–S126.
- OpenAI (2021) CLIP: connecting text and images. Available at: https://openai.com/research/clip (accessed 19 March 2024).
- Parikka J (2023) *Operational Images: From the Visual to the Invisual*. Minneapolis, MN: University of Minnesota Press.
- Parks L and Starosielski N (eds) (2017) *Signal Traffic: Critical Studies of Media Infrastructures*. Urbana, IL: University of Illinois Press.
- Pasquale F (2015) The Black Box Society: The Secret Algorithms that Control Money and Information. Cambridge, MA: Harvard University Press.
- Peters U (2022) Algorithmic political bias in artificial intelligence systems. *Philosophy & Technology* 35(2): 25.
- Prandini R (2015) Relational sociology: a well-defined sociological paradigm or a challenging 'relational turn' in sociology? *International Review of Sociology* 25(1): 1–14.
- Pye, LW and Verba S (2015) *Political culture and political development*. Princeton, NJ: Princeton University Press.
- Radford A, Kim JW, Hallacy C, et al. (2021) Learning transferable visual models from natural language supervision. *arXiv:2103.00020. arXiv.* Available at: http://arxiv.org/abs/2103.00020 (accessed 8 August 2024).

Ranciere J (2013) The Politics of Aesthetics: The Distribution of the Sensible. London: Continuum.

- Rodríguez-Ferrándiz, Raúl. 2014. "Culture Industries in a Postindustrial Age: Entertainment, Leisure, Creativity, Design." *Critical Studies in Media Communication* 31(4): 327–41. https://doi.org/10.1080/15295036.2013.840388
- Rozado D (2023) The political biases of ChatGPT. Social Sciences 12(3): 1-8.
- Seaver N (2019) Captivating algorithms: recommender systems as traps. *Journal of Material Culture* 24(4): 421–436.

- Seaver N (2022) Computing Taste: Algorithms and the Makers of Music Recommendation. Chicago, IL: University of Chicago Press.
- Seyfert R and Roberge J (2016) *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies.* Abingdon: Routledge.
- Staeheli U (2012) Listing the global: dis/connectivity beyond representation? *Distinktion: Journal* of Social Theory 13(3): 233–246.
- Striphas T (2015) Algorithmic culture. European Journal of Cultural Studies 18(4-5): 395-412.
- Striphas T (2016) Culture. In: Peters B (ed.) Digital Keywords: A Vocabulary of Information Society and Culture. Princeton, NJ: Princeton University Press, pp. 70–80.
- Swedberg R (2020) Exploratory research. In: Elman C, Mahoney J and Gerring J (eds) The Production of Knowledge: Enhancing Progress in Social Science. Strategies for Social Inquiry. Cambridge: Cambridge University Press, pp. 17–41.
- Swist T, Humphry J and Gulson KN (2023) Pedagogic encounters with algorithmic system controversies: a toolkit for democratising technology. *Learning, Media and Technology* 48(2): 226–239.
- Tiu E (2020) Understanding latent space in machine learning. Available at: https://medium. com/data-science/understanding-latent-space-in-machine-learning-de5a7c687d8d (accessed 6 August 2024).
- Wasik B (2009) And then there's this: How Stories Live and Die in Viral Culture. London: Penguin.
- Weber J (2016) Keep adding. On kill lists, drone warfare and the politics of databases. *Environment and Planning D: Society and Space* 34(1): 107–125.
- Werbin KC (2017) The List Serves. Population Control and Power. Amsterdam: Institute of Network Cultures.
- Wikipedia (2024) List of Political Ideologies. *Wikipedia*. https://en.wikipedia.org/w/index. php?title=List of political ideologies&oldid=1286014186
- Williams R (1958) Culture is ordinary. In: Szeman I and Kaposy T (eds) Cultural Theory: An Anthology. London: Wiley, pp. 53–59.
- Williams R (1976) *Keywords: A Vocabulary of Culture and Society*. New York: Oxford University Press.
- Witherspoon G (1980) Language in culture and culture in language. *International Journal of American Linguistics* 46(1): 1–13.
- Wright S (1998) The politicization of 'culture'. Anthropology Today 14(1): 7–15.
- Xu H, Xie S, Tan XE, et al. (2024) Demystifying CLIP data. *arXiv:2309.16671. arXiv.* Available at: http://arxiv.org/abs/2309.16671 (accessed 23 June 2024).
- Young LC (2017) *List Cultures: Knowledge and Poetics from Mesopotamia to BuzzFeed.* Amsterdam: Amsterdam University Press.

Author biographies

Luke Munn is a Research Fellow at the Centre for Digital Cultures & Societies at the University of Queensland, Australia. The author of six books and numerous articles, his work combines digital methods and critical insights from across the humanities to explore the social, political, and environmental implications of contemporary technologies.

Adarsh Badri is a PhD student at the School of Political Science and International Studies (POLSIS), University of Queensland, Australia. Earlier, he was a commissioning editor at *E-International Relations*. His research interests include identity, emotions, postcolonialism, IR theory, and foreign policy, focusing on South Asia.