New College of the Humanities

British Wittgenstein Society Conference

Symposiast Abstracts

Wittgenstein and AI

New College of the Humanities

Wittgenstein: Philological and Semantic

Alois Pichler (University of Bergen)

29 July 10:30

In my talk, I will first briefly present the web resources and tools offered by the Wittgenstein Archives at the University of Bergen (WAB) for research on Wittgenstein. They include static editions of Wittgenstein primary sources on Wittgenstein Source http://www.wittgensteinsource.org/, dynamic and interactive editions of the Wittgenstein Nachlass on Wittgenstein IDP http://wab.uib.no/transform/wab.php?modus=opsjoner, faceted search and browsing of Wittgenstein metadata on Wittgenstein SFB http://wab.uib.no/sfb/ and lemmatized and other advanced text search on Wittgenstein WiTTFind http://wittfind.cis.uni-muenchen.de/. Then I will present WAB's work on a comprehensive computational ontology for capturing the semantics of the Wittgenstein domain (see http://ubbdev.gitlab.io/wab-ontology/index-en.html and http://wab.uib.no/cost-a32_philospace/wittgenstein.owl). Finally, I will discuss some of the possibilities and challenges that arise from "ontologizing" the truly philosophical parts of the ontology's Subject class, namely its subclasses Concept, Claim and Argument.

New College of the Humanities

Distributional Semantics and Computational Methods for Tracing Words' Meaning Across Time

Barbara McGillivray (King's College London)

29 July 11:10

Computational linguistics develops algorithms to analyse language data automatically to shed new light into linguistic phenomena. The state-of-the-art approaches to computational modelling of words' semantics rely on the principles of Distributional Semantics, according to which meaning arises from the way words are used in real-world language utterances. Distributional models have evolved from Wittgenstein's later work, but also from the work of Zellig Harris and J.R. Firth, among others. The ability of distributional semantics to fully explain word meaning is object of debate, but there is no doubt that it has had a very strong success in computational research. With our ever-increasing access to large textual collections, whether digitised or born-digital, its principles can be readily operationalised and employed in algorithms. Unlike rule-based systems (such as part-of-speech taggers that use hard-coded rules to detect categories of words and apply this knowledge to further tasks), distributional methods have the ability to learn directly from textual examples. In this talk I will present my research on developing computational models for lexical semantic change detection and analysis in historical texts using distributional approaches. Over time, new words enter the language, others become obsolete, and existing words acquire new meanings. In Old English thing meant 'a public assembly' and now means more generally 'entity'; chill originally meant 'to cool' and has metaphorically been extended to 'to relax'. I will share my experience of working at different scales and in a range of interdisciplinary projects, from Ancient Greek and Latin to web archives, Twitter and emoji, and will discuss the challenges and opportunities of working in interdisciplinary projects in Data Science, Digital Humanities and Computational Linguistics.

New College of the Humanities

Computational Philosophy: Reflections on the PolyGraphs Project

Brian Ball (New College of the Humanities)

29 July 11:50

In this paper, we situate computational philosophy relative to other digital humanities and computational social science practices, based on reflections stemming from our research on the PolyGraphs project in social epistemology. We begin by describing PolyGraphs. An interdisciplinary project funded by the Academies (BA, RS, and RAEng) and the Leverhulme Trust, it uses philosophical simulations (Mayo-Wilson and Zollman, 2021) to study how ignorance prevails in networks of inquiring rational agents. We deploy models developed in economics (Bala and Goyal, 1998), and refined in philosophy (Zollman, 2007; O'Connor and Weatherall, 2018), to simulate communities of agents engaged in inquiry, who generate evidence relevant to the topic of their investigation and share it with their neighbours, updating their beliefs on the evidence available to them. We report some novel results surrounding the prevalence of ignorance in such networks, focussing on the relative importance of various structural network properties - e.g. network size, density, diameter, and related patterns - in determining this. In the second part of the paper, we compare our own to other related practices. We begin by noting that, in certain sorts of digital humanities projects, the computational component does not directly support the humanities research itself so much as its presentation. In such projects, the digital and the humanities are simply grafted together, not fully intertwined and integrated. PolyGraphs is notably different: the computational work directly supports the investigation of the primary research questions, which themselves belong decidedly within the humanities in general, and philosophy in particular. This suggests an affinity with certain projects in the computational social sciences. But despite these real similarities, there are differences once again: the computational philosophy we practice aims not so much at description and prediction as on answering the normative and interpretive questions that are distinctive of humanities research.

New College of the Humanities

The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law

Sandra Wachter (Oxford)

29 July 14:00

Artificial intelligence is increasingly used to make life-changing decisions, including about who is successful with their job application and who gets into university. To do this, AI often creates groups that haven't previously been used by humans. Many of these groups are not covered by non-discrimination law (e.g., 'dog owners' or 'sad teens'), and some of them are even incomprehensible to humans (e.g., people classified by how fast they scroll through a page or by which browser they use). This is important because decisions based on algorithmic groups can be harmful. If a loan applicant scrolls through the page quickly or uses only lower caps when filling out the form, their application is more likely to be rejected. If a job applicant uses browsers such as Microsoft Explorer or Safari instead of Chrome or Firefox, they are less likely to be successful. Non-discrimination law aims to protect against similar types of harms, such as equal access to employment, goods, and services, but has never protected "fast scrollers" or "Safari users". Granting these algorithmic groups protection will be challenging because historically the European Court of Justice has remained reluctant to extend the law to cover new groups. This paper argues that algorithmic groups should be protected by non-discrimination law and shows how this could be achieved.

New College of the Humanities

AI and the Ethics and Aesthetics of Apologising

Burkhard Schafer (University of Edinburgh)

29 July 14:40

When human decision makers fail in their legal duties, we often expect, and sometimes demand in law, that they apologise. If machines that replaced the human fail, does it make sense to make the same requirement of them? The talk will look at apologies as a regulatory tool and its applicability to AI though a Wittgensteinian perspective.

New College of the Humanities

Data Handling Online: Rethinking the Data Control Assumption

Stephen Dness (New College of the Humanities)

29 July 15:20

There is an active debate about the role of information gathering and sharing online. Websites and advertisers can track large quantities of information, yet users continue to use the websites. A typical example would be seeing a tailored advert on a newspaper, and then seeing a similar advert on a different website, prompted by search history. This presents a puzzle: do consumers not mind? Or do they simply not know? If they did know, would they mind? It probably depends on what is being done with the information. There is a difference between an invasive data gathering practice (e.g., searches for divorce lawyers; searches for health condition) and more innocuous ones (e.g. searches for new shoes). As the systems involved are new, legal and regulatory concepts have struggled to keep pace. There are rudimentary consumer protections in the GDPR and its UK retention (the Data Protection Act), but these systems were not designed for some data handling uses. The major issue in GDPR is that it speaks only to "personal data," applying a jurisdictional test of identifiability. This means that "pseudonymous" data (User 123) is not necessarily covered. Yet some concerns arise even with pseudonymous data. For example, someone else might see adverts for a divorce lawyer when sharing a PC. In this example, identity is not directly revealed to the advertiser, and thus the data handling is arguably not regulated, since only "User123" is involved, rather than anyone's identity per se. The difficult question is: when do the cross correlations between adverts and the associated profiling practices become a problem, even if identity is not revealed? There is very little if any attention to this in current legal and regulatory debates. Good examples include the UK Information Commissioner's Office's November 2021 AdTech Opinion, which skirts this issue (pp.27-30), and the Belgian Data Protection Authority's "TCF" cookies decision, which similarly skirts the issue of defining harm from data handling. Instead, the documents proceed on the somewhat dated assumption that consumers should have control over data – sensible as an aim, but does it work if systems are so complex that only a PhD statistician could fully understand them? A concept of harm is essential to meaningful cost-benefit analysis to ensure that consumers benefit. At the same time, personalisation systems can confer significant benefits, including in advertising, although the point is of broader application. A tailored advert may be helpful to the consumer, and will provide a content producer with much more revenue. Studies estimate this to be between 70% and 500% more effective: shoes advert to those looking for shoes, etc. This helps fund free content on the internet and has an important democratic function in ensuring free access to the press by avoiding paywalls, and allowing other ad-funded models (Spotify) enhancing consumer access, especially to those with limited funds for subscriptions. Tailoring is also important to green economy transitions, by allowing green brands to target those most keen to switch to them, and to help determine ways to appeal to customers who are not currently switching. There is a risk of underestimating the benefits of responsible data handling if it is simply banned outright. Additionally, large technology companies with large systems (Google, Apple, Facebook) may have an incentive to prevent data handling by lobbying for restrictive rules, even where data use is innocuous. This can undermine competition from small and medium sized businesses. The paper considers this developing landscape and, specifically, the shortcomings in the consent-based model which currently prevails, contrasting them with alternatives based on better-defined, and more tailored, concept of online harm.

New College of the Humanities

The Later Wittgenstein on Idealization and Abstraction in Logic

Oskari Kuusela (University of East Anglia)

29 July 16:15

In my talk I outline Wittgenstein's later account of idealization and abstraction in logic. In its original context this account serves purposes such as the following. 1) It explains how in the study of logic it is possible to describe uses of language in simple and exact terms without misleadingly simplifying and falsifying the targeted uses, when it is acknowledged that the latter are complex and vague. Thus the account helps to explain how logic can meet its standards of rigour when the legitimacy postulation of simple and exact logical structures underlying the surface of language, such as Wittgenstein postulated in the Tractatus, is questioned. 2) Wittgenstein's account helps to solve a confusion about ideal notions of logic which he fell into in the Tractatus believing that, since everyday language does not meet the standards of simplicity and exactness of logic, it is necessary to postulate simple and exact rules of language use hidden under the surface of language that do meet those standards. By contrast, the later Wittgenstein maintains that rather than being a feature of the uses of language targeted for logical clarification, ideal features, such as simplicity, exactness, and exceptionless logical generality and necessity, are to be recognized features of the modes of representation of logic. As regards the study and development of AI, Wittgenstein's account of idealization and abstraction in logic helps to think about how it is possible to manage complex information in simple and exact terms, without representing this information in false simplicity, and with the possibility of organizing and re-organizing this information in a variety of ways in response to different interests and concerns.

New College of the Humanities

Artificial Argumentation: From Logic to Language and Back

Serena Villata (University of Nice)

29 July 16:55

Argumentation is the process by which arguments are constructed and handled. In formal computational models of argument, each argument is a set of assumptions that, together with a conclusion, is obtained by a reasoning process. Two kinds of views on argumentation can be highlighted, namely monological and dialogical. In the former, a single agent has the knowledge to construct arguments to support and attack a conclusion to reach a decision, while, in the latter, a group of agents interacts to construct arguments supporting or attacking a particular claim and finally deliberate. Although natural language argumentation has attracted the attention of philosophers and rhetoricians since Greek antiquity, it is only very recently that the methods and techniques of computational linguistics and machine learning have become sufficiently mature to tackle this extremely challenging topic. Argument mining, the new and rapidly growing area of Natural Language Processing (NLP) and computational models of argument, aims at the automatic recognition of argument structures in large resources of natural language texts. In this talk, I will provide an overall view of Artificial Argumentation highlighting the role of logic and language in this field, with a focus on two timely research lines, namely fallacious argumentation and argumentation quality.

New College of the Humanities

Modelling Analogical Reasoning: One-Size-Fits-All?

Ioannis Votsis (New College of the Humanities)

29 July 17:35

One of the most common forms of reasoning in science is reasoning by analogy. Roughly speaking, such reasoning involves the transposition of solutions that work well in one domain to another on the basis of analogous features between the two domains. Sometimes such reasoning works (e.g. artificial and natural selection) and sometimes it doesn't (e.g. Vulcan and Neptune). Two general reactions to the problem of modelling the logic of analogical reasoning have emerged as a result: There are those who attempt to construct increasingly complex but still universal models of analogical reasoning in order to better discriminate between cases where it works and cases where it doesn't (Bartha 2010; Hesse 1966). And there are those who give up on the universal model approach and argue in favour of localised models (Norton 2021). In this talk, we assess the merits of each approach in the context of the Wittgensteinian family resemblance conception of scientific categories. Moreover, we assess the impact of computational attempts to articulate and operationalise analogical reasoning, particularly in the field of Artificial Intelligence (Prade and Richard 2014), on the debate between universalists and localists.

New College of the Humanities

Wittgenstein and Turing on AI: Myth Versus Reality

Diane Proudfoot (University of Canterbury, New Zealand)

30 July 10:30

A standard view of Wittgenstein and Turing is that both were philosophical behaviourists regarding the mind. Amongst commentators sympathetic to Wittgenstein, a different view is usual—namely, Turing was a behaviourist, and Wittgenstein was Turing's first critic. Wittgenstein, it is claimed, posed 'criticisms of Turing's computational thinking theory', 'critiques of Turing's idea that people are computers', and 'counter-arguments' to Turing's 'ideas of intelligent machinery' (Saariluoma and Rauterberg 2015, 2016; Taylor et al. 2010). Wittgenstein, many scholars agree, had 'profound doubts and open disagreements with Turing' on the possibility of artificial intelligence (Liu 2021). Proponents of the alternative sympathetic view align Wittgenstein with AI naysayers. Wittgenstein held, Hacker (1996) claims, that 'it makes no sense to ascribe thought or thoughtlessness, understanding, misunderstanding or failure to understand to machines'. For Wittgenstein, the question Could a machine think? is 'logically absurd' and 'conceptually ill-formed' (Shanker 1998). In addition, it is declared, Wittgenstein would have rejected Turing's computer-imitates-human game as a test of intelligence in machines; in his view, any such attempt 'must be doomed to failure' (Taylor 2015). I shall argue that both the standard and the sympathetic views are myths. Neither Turing nor Wittgenstein was a behaviourist, and there are intriguing similarities in their approaches to the possibility of AI.

New College of the Humanities

Alan Turing – Accidental Codebreaker?

David Kenyon, Research Historian, Bletchley Park Trust

30 July 11:10

When the name Turing is mentioned, the popular response is to identify the Second World War Codebreaker. Alan Turing is rightly famous for his contribution to the breaking of wireless messages encrypted using the German armed forces' Enigma cipher machine. As part of a research team headed by Dillwyn Knox at Bletchley Park he was responsible for the design of the bombe machines, used in key-finding for Enigma, and the resulting contribution to Allied success in the war is well known. He also worked on other ciphers including the Lorenz teleprinter system, as well as research into speech encryption. His sexuality, and his early death at the age of only 41 have also made him something of a folk hero and diversity icon.

However Turing is also remembered as a founding father of the modern digital age, due to his fundamental theoretical work on mathematical logic, and what would later become computer science and artificial intelligence. Arguably had he lived to continue a full career he would no doubt be much better remembered for his mathematical and computational work than for his codebreaking. Inspired by the contents of one of his wartime notebooks, this presentation considers to what extent Turing was ever a serious cryptanalyst, following the model of other WW2 cryptographers such as William Friedman or John Tiltman, or whether Turing's work on codes and ciphers was a side-line, and even a distraction from his true vocation as a mathematician and theorist. Biographers and film makers have seized upon Turing's pre-war encounters with codes and ciphers, in order to argue that he was somehow destined to be a cryptanalyst. However these encounters can be viewed just as readily as coincidences, or amusements, used as tools in the development of his real interests. By re-examining what he did during his time with the Government Code and Cypher School (GC & CS, now GCHQ) this presentation seeks to ask; was Alan Turing only accidentally a codebreaker? Did he really care for the work? Was breaking Enigma a defining moment in his career, or merely a temporary wartime diversion from a higher purpose?

New College of the Humanities

Wittgenstein and Turing's Discussion of Contradiction and Inconsistency

Ásgeir Berg Matthíasson (University of Iceland)

30 July 11:50

In this talk, I give a close reading of Wittgenstein's discussion of inconsistency and contradiction, mostly centering on his discussion with Alan Turing in the Lectures on the Foundations of Mathematics. I focus especially on the so-called 'falling bridges'-objection given by Turing. I will argue that Wittgenstein's position is that if contradictions arise in some practice of language, they are not necessarily fatal to that practice nor necessitate a revision of that practice. Turing's objection is that if we are engaged in a practical matter, for example, building a bridge, an inconsistent calculus could lead to catastrophe, since anything follows from a contradiction. As is well-known in the literature, Wittgenstein had recommended that if we are worried about inconsistency leading to triviality, we can always refuse to draw conclusions from the contradiction, for example by abolishing the ex falso quodlibet rule. With that in the background, Wittgenstein's answer to Turing is that if we run into trouble building our bridge, it is either because we have made a calculation mistake (drawn a conclusion from the contradiction, which is not allowed by the new rules of our calculus) or our calculus does not actually describe the phenomenon it is intended to model. The possibility of either kind of error is not particular to contradictions nor inconsistency, and thus contradictions do not have any special status as a thing to be avoided in Wittgenstein's view. I explain why, in Wittgenstein's view, there is no third option. I then briefly examine the idea of 'naturalness' that seems to underlie Wittgenstein's position and allows him to say that an inconsistent practice is not necessarily therefore impractical or defective. In the Lectures, this notion seems to play a constitutive role for the correctness conditions of concepts and plays a crucial role in Wittgenstein's discussion of rule-following there. He repeatedly seems to argue that we 'extend' our practices in the way we find most natural and then that way is adopted as the standard—of applying the concept correctly. The idea then, it seems to be, is that an inconsistent concept can be perfectly usable as such and require no revision, for example where it fails only in edge cases that have no practical implications, even in theory (as Wittgenstein's discussion of the liar paradox seems to entail) or where the inconsistency is such that it simply will never be come manifest in practice. In such cases, avoiding paradox is more akin to an aesthetic choice than a practical or epistemological necessity.

New College of the Humanities

Management, Automation, and the Consequences of Denial

Paul Standish (UCL IOE)

30 July 14:00

In the 1950s BF Skinner expressed his belief in the potential of teaching machines for revolutionising education. The machines he referred to led students through a step-by-step process: short passages of exposition and explanation were followed by a test activity; success in a test led the student to the next step, while failure provided them with some further explanation at the same level and a further test. This process became known as "programmed learning". Learners could proceed at their own pace, with frequent reinforcement of correct behaviour. If today I want to learn how to use spreadsheets, I can find programs on the Internet that will lead me through a similar step-by-step process. In the 1980s there was a sense of urgency about the need to bring computers into schools, and courses in computer programming became widely available in secondary education, usually as alternatives to more familiar curriculum subjects. But this practice waned with the realisation that it was not so much computer programming that was needed as an essential skill for the future but competence with computer applications. Meanwhile, more computerisation had taken effect at the less visible level of administration. This encompassed changes in forms of assessment and monitoring, in which the gathering of data in numerical form was increasingly required. What gets measured gets managed, and what is measured easily is easy to automate. The advent of the internet dramatically expanded the potential importance of this new technology's influence on education. It answered to the principled aim of extending education to those who had previously been in some degree excluded. In post-compulsory education especially, there was the new ideal of "open learning": courses would be available when and where students could access them, with each customised to the student's individual needs – potentially on a global scale. Such emancipatory ambitions were compromised, however, by a more managerialist impulse, and "lifelong learning" came to be synonymous with an agenda of upskilling for the knowledge economy. Analysis of the emergence of AI in education needs to take this background into account, as well as its policy-rhetorical context. Here is a sample from amongst the true believers: AI's emergence in education must reinvigorate "the deeper learning goals of a modern education: Versatility, for robustness to face life and work; Relevance, for applicability, and student motivation; Transfer, for broad future actionability" (Holmes, Bialik, and Fadel, Artificial Intelligence In Education, 2019). This increasingly familiar educational newspeak comes with a rather stark separation of content and method, where much of the focus is on the latter, especially on learning and teaching (or "delivery"). The rhetoric has been a feature of the developments described above, with technology positioned at the forefront of "school effectiveness and improvement". Against this background, my discussion considers three Wittgensteinian lines of critique, regarding (1) the reduction of knowledge content and the displacement of judgement, (2) the dominance of algorithmic reasoning; and (3) the advent of the robot teacher.

New College of the Humanities

What Happens (and Doesn't Happen) When AI Teaches Grownups?

Chiara Alfano (New College of the Humanities)

30 July 14:40

Picture a VLE-integrated chatbot. Its algorithm allows it to scan student questions for certain key words – let's say "block," "pillar," "slab" and "beam." When it detects one of these words in a student question, it returns a predetermined answer. Though an efficient transmission of information, we would hesitate to call even a sophisticated version of this teaching. Or would we? My hypothetical chatbot is inspired by the builders who appear at the very beginning of the Philosophical Investigations, where A calls out material he needs, and B passes them to him. This type of communication is, Wittgenstein is quick to point out, not fit for teaching children "everything we call language" (PI, § 3). And although Wittgenstein imagines the builders in the context of thinking about language as forms of life, the alacrity of his dismissal has undoubtedly also to do with the fact that the scene does not go anywhere near capturing what happens when we teach. The clue for the absurdity of this picture lies in Wittgenstein's use of Abrichtung (training) rather than Beibringen or Lehren (teaching). What isn't happening when my hypothetical chatbot (the builders' assistant B, in Wittgenstein's example) returns answers to the student (builder A)? Stanley Cavell is finely attuned to the dystopian connotations this picture implies for the teaching of children. Imagining the builders' hypothetical children, he finds that they would "have to exert great efforts to suppress certain natural responses of the children" (PP, 164), adding that he could "imagine robots, or men hypnotized, doing the things the builders do at the same four calls." (146). This vision does not allow for the "natural responses" of and, indeed to the children. Simply put, forms of life aren't happening. Cavell's reading of this passage emphasises not only that Wittgenstein is here preoccupied with how language as forms of life is taught but that forms of life are irrevocably part of what happens when we teach. We are reluctant to let go of this aspect of teaching the younger the child in question – but what about the teaching of grownups? In this talk, I will consider the use of adaptive learning or intelligent tutor systems (which personalise the learning experience by altering the pace, order or level of the learning) in Higher Education to think about what happens and doesn't happen when AI teaches grownups. I will approach this question by way of Cora Diamond's "The Difficulty of Reality and the Difficult of Philosophy" (2003) in which she draws on a dimension of teaching that is often overlooked, particularly in the context of HE, to ask: can AI-powered teaching be not only be a preparation but also part of the moral life? Can AI – like Diamond's good teacher – be responsive and awake to the whole student and hold space for the possibility of human difficulty this entails? Can AI-led teaching be adventure and improvisation? Can we picture AI that would be able to do this? If yes, what might it look like? If not, why not? And why is it worth for us human HE educators to ask these questions?

New College of the Humanities

Artificial Intelligence: Lived Experience

Mark Martin (New College of the Humanities)

30 July 15:20

Talk: Mark Martin MBE will explore the rise of Ai in edtech and to what extent does the technology consider the lived experience of the individual. He will explore case studies where Ai has been used in education and the impact it has had on the users.

Background:

Artificial Intelligence is now starting to be used in all areas of our lives, its decision making process has the ability to determine our fate based on datasets and algorithms. The reliability of the data and individuals programming the datasets could be questioned to see whether their lived experience reflects the performance of the product. Also does the product take into consideration the user's lived experience and circumstance.

The A-level debacle within the UK highlighted the flaws Ai had when trying to predict the outcomes for students based on their GCSE grades. The Ai system confirmed the stereotypes that exist within the education system. Some may argue it wasn't a flaw, it amplifies the biased views that are in the world today and its perception of the 'other'.

As we start to mass adopt AI in education technology known as edtech, how do we prevent further biased views creeping in the technology and output for teachers and students?

New College of the Humanities

Automated Misinformation Detection Using Natural Language Processing

Elena Kochina (Queen Mary University/ Alan Turing Institute)

30 July 16:15

Online misinformation had a dramatic impact on our society in recent years, affecting people in many sectors, from finance, health to politics. The importance of combating the spread of fake news has been acknowledged by the public, companies and governments. The problem is being addressed through policies, journalistic work, media literacy of the population, social media platforms response and research. While various initiatives have been launched to address this problem, manual fact checking and verification cannot scale to address the amount of unverified information circulating and cannot be easily performed in real-time. There is therefore increasing interest and need in automated Machine Learning-based methods to assist with the verification. Our research aims to tackle this problem using Machine Learning and Natural Language Processing methods to propose systems for automated rumour verification that would assist journalists in their work. In this talk I will discuss various types of existing approaches to automated rumour verification, overview recent advances and outline open challenges that rumour verification models are facing, and share my view on how to tackle them.

New College of the Humanities

Authenticity and Trust Online

Elmar Unnsteinsson (University College Dublin)

30 July 16:55

Online language games have introduced new moves into our communicative repertoire. I will focus specifically on links and hashtags. I propose a semantics according to which they denote noncentral speech acts, like 'moreover,' and 'however.' Links are standardly used to provide implicit epistemic support and can easily be used to manipulate the audience. Manipulative hashtags are more difficult to pull off because they normally involve minor changes in the text of the hashtag, to lead the audience to a misleading source. I argue that a theory of the sincerity conditions of online noncentral speech acts can help us to identify potential repair strategies for rampant disinformation on popular social media. Very roughly, hashtags or links should ideally give access to time-ordered contributions to a conversation aimed at answering related questions. The web now records multiple occasions on which someone has answered the question 'How many people died at Pearl Harbor?' Almost every source will say 2,403 and exclude the number of Japanese killed. Could we somehow represent a list of occasions when this question was answered and propose a convention whereby any new answer-occurrence is linked to the list? Perhaps. Regardless, I argue that it is high time to creatively imagine different and better online communicative practices, because the current disinformation crisis is largely due to our current lack of reliable conventions to signal sincerity, trustworthiness, or authenticity in online exchanges.

New College of the Humanities

Wittgenstein, Distributional Semantics and Large Language Models

Fintan Mallory (University of Oslo)

30 July 17:35

Current research on language modelling, in particular, work grounded in the tradition of distributional semantics, often cites the later Wittgenstein as an inspiration. The 'distributional hypothesis' which underlies this framework is proposed as an implementation of the Wittgensteinian idea that 'the meaning of a word is its use in the language' (PI 43). The hypothesis builds on Frege's Context Principle, the idea that an expression only has meaning in the context of a proposition, and proposes to model meanings by their contexts. A 'meaning' is modelled as the set of contexts in which a word has occurred within some data set. These contexts can be represented as multidimensional vectors within a meaning space such that words with similar vectors, according to metrics like cosine similarity, have similar meanings. This provides representations of meaning that are amenable to computational processing, which is a technological virtue, as computers are much better at adding and subtracting vectors than 'grasping senses'. The shift from developing encyclopaedic entries for expressions, to the modelling of distributional contexts, has lead to a massive leap forward in a range of natural language processing tasks, from translation and document analysis to text generation. Furthermore, while earlier distributional models of meaning struggled to capture polysemy and homonymy, more recent context-sensitive transformer-based models do a much better job at generalising, or grasping the 'complicated network of similarities' that bind uses into a family. Unfortunately, this also leads to a problem. While earlier models represented contexts in a transparent fashion, so that anyone could work out the contextual representation of a word by counting how often a word co-occurs with a context, more recent models are not deterministically derived from a dataset but produced by neural networks trained on the dataset to solve prediction tasks. The result is denser, more successful, but less interpretable representations. In particular, it isn't clear what the dimensions of the representation represent. In this talk, I will examine the relevance of private-language-like considerations for these language models, specifically trying to identify if the difficulty we face interpreting these models should caution us from claiming that they represent meanings.

New College of the Humanities

From `Meaning is Use' to Unsupervised Grounding of Color Terms

Anders Søgaard (University of Copenhagen)

31 July 10:30

I present new evidence that pretrained language models - part of many artificial intelligence technologies - embed color terms in a way that is near-isomorphic to our perceptual color space. This is remarkable, since pretrained language models are trained on distributional information alone, i.e., higher-order co-occurrence patterns. The same holds for other domains, including direction words. If pretrained language models learn representations that are near-isomorphic to perceptual and physical spaces, we can easily ground these subspaces, even in the absence of explicit supervision. This challenges the common view that computers, being confined to syntactic symbol manipulations, can never acquire meaning (Searle, 1980). Moreover, if distributional evidence is sufficient to learn intentional semantics, does this open for alternative theories of language evolution and acquisition?

New College of the Humanities

Cultural differences in the cognitive aspects of colour geometry

Dimitries Mylonas (NCH), Maria Federica Norelli (NCH), Julies Davidoff (Goldsmiths University of London)

31 July 11:10

In his Remarks on Colour (1951/1977) Wittgenstein asked: “Can’t we imagine certain people having a different geometry of colour than we do? That, of course, means: Can’t we imagine people having colour concepts other than ours?” He also famously argues for the impossibility of colour concepts such as reddish-green and bluish-yellow which he equates with the feeling of a southwestern north wind. In this study, we compare geometric features of colour concepts derived from experimental colour naming data in English and in Himba (a language of a remote culture from northwest Namibia), and discovered using methods from machine learning. We find cultural differences between English and Himba speakers in their number of colour names and in the size, shape, and location of their corresponding colour categories. Himba might even legitimately be said to understand reddish-green and bluish-yellow while English might not. These results show that we can empirically determine the relationship between different colour concepts within colour space to facilitate colour communication within and across languages.

New College of the Humanities

Imaginative attempts to LW´s Aspects in Language, AI-sentiment Investigations-Approach

Tatjana Christelbauer (Agency for Cultural Diplomacy, Vienna)

31 July 11:50

The Gedankenexperiment ^thought experiment is a part of the multimedia artworks which has been created on interface of visual arts with applied linguistics, cognitive humanities and sociology in response to some of the currently most urgent societal challenges: hate speech, physical violence with a therapeutic approach to prevention of violence, leaning on Ludwig Wittgenstein's analyses about imagination (Philosophische Untersuchungen, Language Games, Blue & Brown Notebooks, Zettel), Ludwig Wittgenstein - Lecture on Ethics, J. P. Lederach notions on Moral Imagination, and the B. Brecht's Radio theory, assistive technologies, family resemblance of colour by Orange Featherflight guided by following questions for art-based investigations:

Ethical sentiments in AI: can speech acts be re-imagined by seeing in aspects as provided AI and the use of assistive technologies for screen readers?
The thought experiment ^Gedankenexperiment **Orange Feather**:

See ^imagine the Orange Feather over there?
Imagine the sound of the Orange Feather flying.
Now, translate it into German, Italian, Croatian, ...
(not the word, but the object-as for screen readers)

Imagination also fulfills a methodical role, because we can visualise imaginary cases similar to Gedankenexperimente (LW. in PU, §312)

The result of the imaginative experiment will be so clear that there will be no need to conduct a physical experiment at all (...) "If fear is frightful and if while it goes on, I am conscious of my breathing and of a tension in my face,
is that to say I feel these feelings frightful? Might they not even be a mitigation?"
(L.W. Zettel, 484-499)

"One knows the position of one's limbs and their movements... [with] no local sign about the sensation. (...) such attention to one's movements and feelings can hinder the smooth execution of willed action (L.W. Zettel, 483).