



HAL
open science

AI is coming and our languages need to catch the wave

Mélanie Jouitteau

► **To cite this version:**

Mélanie Jouitteau. AI is coming and our languages need to catch the wave. Gareth Watkins. Language and Technology in Wales: Volume II, II, 2024, 978 1 84220 207 4. hal-04793230

HAL Id: hal-04793230

<https://hal.science/hal-04793230v1>

Submitted on 20 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Language and Technology in Wales: Volume II

Watkins, Gareth; Prys, Delyth; Prys, Gruff; Jones, Dewi; Cooper, Sarah; Williams, Meinir; Vangberg, Preben; Ghazzali, Stefano; Gruffydd, Ianto; Farhat, Leena; Grobol, Loïc; Jouitteau, Mélanie; Morris, Jonathan; Ezeani, Ignatius; Young, Katharine; Davies, Lynne; El-Haj, Mahmoud; Knight, Dawn; Jarvis, Colin; Barnes, Emily

Published: 01/11/2024

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Watkins, G. (Ed.), Prys, D., Prys, G., Jones, D., Cooper, S., Williams, M., Vangberg, P., Ghazzali, S., Gruffydd, I., Farhat, L., Grobol, L., Jouitteau, M., Morris, J., Ezeani, I., Young, K., Davies, L., El-Haj, M., Knight, D., Jarvis, C., & Barnes, E. (2024). *Language and Technology in Wales: Volume II*. (1 ed.) Prifysgol Cymru Bangor.

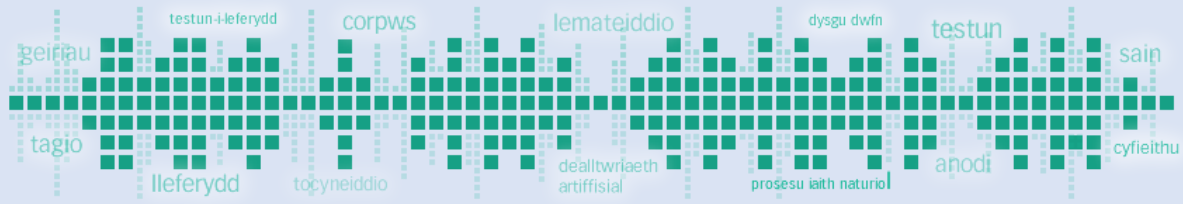
Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Language and Technology in Wales: Volume II

Editor: Gareth Watkins



PRIFYSGOL
BANGOR
UNIVERSITY



This ebook was first published in 2024 by
Bangor University, College Road, Bangor, Gwynedd LL57 2DG
www.bangor.ac.uk

International Book Number (ebook):

ISBN 978-1 84220-207-4.

The text has been released under the Creative Commons BY 4.0 license
<https://creativecommons.org/licenses/by/4.0/>, which allows you to reuse and modify it in
any way if you provide appropriate acknowledgment. See license text
<https://creativecommons.org/licenses/by/4.0/> for more details.

Design and proofreading assistance from Prof. Delyth Prys and Stefano Ghazzali. This book is
also available in Welsh under the title *Iaith a Thechnoleg yng Nghymru: Cyfrol II*, ISBN number
978-1 84220-208-1.

Addendum

MÉLANIE JOUITTEAU - L'UNIVERSITÉ BORDEAUX MONTAIGNE AND L'UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

AI is coming and our languages need to catch the wave

My point is simple. Our languages must become AI compatible if they are to survive the coming deployment of applications based on personal data.

Artificial Intelligence (AI), a system characterized by models that are trained using extensive datasets, including linguistic data, has become an integral part of our daily lives. This is evident in our frequent interactions with predictive typing and voice recognition technologies on smartphones. Currently, we are witnessing only the initial phase of its societal integration. With this context in mind, I draw a parallel between the social integration of AI and that of mobile phones. In a detailed analysis comprising twelve points, I outline the manner in which the previous technological revolution, namely mobile phones, became embedded in our lifestyles. I draw comparisons with the ongoing AI revolution at each of these stages.

The present phases in AI development are laying the groundwork for the implementation of applications reliant on personal data. I offer a detailed explanation of why, from my perspective, AI applications that utilize personal data pose a significant risk of drastically reducing the world's linguistic diversity in the near future. Although the swift integration of AI into our lives seems almost inevitable, the specific threat of linguistic homogenization can be mitigated. This mitigation is possible through the adaptation of our languages to be compatible with these AI models.

I advocate for a collective effort encompassing public, industrial, scientific, and societal sectors. This mobilization is crucial to ensure that human languages successfully navigate, rather than drown in, the AI tide.

Do you possess a mobile phone?

As a member of Generation X, I witnessed the advent of mobile phones and smartphones. This experience has led me to draw parallels between societal responses to mobile phones 25 years ago and contemporary reactions to AI. It appears to me that society's response to these technological revolutions bears striking similarities. In the 1990s, individuals who were initially hesitant about mobile phones now routinely carry them. I have methodically catalogued the various factors that contributed to this shift in attitude, tracing the technology's societal penetration from its nascent stage to complete integration and points of no return.

Let us reflect on this transition. Those who experienced this shift, particularly those who were initially resistant to mobile phones, can each recall their personal moment of capitulation. Similarly, observers of AI's evolution can discern parallels with the current stage of its development.

In the beginning, it's always...

1 THE AGE OF THE PRIVILEGED FEW

This phase is characterized by the rapid adoption of new technology by a small group of enthusiasts who possess significant economic, cultural, and technological power. Such adoption bestows upon the technology a symbolic status of high social standing. The 'Age of the Privileged Few' continues if the technology remains inaccessible to the majority. In contrast, if the technology becomes widely affordable, it is then described as 'democratizing.' This term is misleading because it bears no democratic value; it merely indicates that the product is no longer exclusive to an elite group. This process does not involve democratic procedures like national elections or referendums to determine the adoption of new technology. Market forces predominantly drive these changes. This phase was unmistakably evident in the case of mobile phones and, since the advent of ChatGPT, has increasingly applied to AI. Notably, ChatGPT offers some of its services for free, unlike mobile phones, further accelerating the spread.

2 PROFESSIONAL REQUIREMENTS

The advancement of technology alleviates numerous professional duties and becomes deeply ingrained in various occupations. The ubiquity of mobile phones has facilitated constant connectivity for work-related tasks, a development that was initially welcomed and eventually became a necessity across many professions, particularly in fields like healthcare where doctors on call must be readily accessible. Especially in an era characterized by widespread unemployment, resistance to adopting mobile technology increasingly results in an exit from these professions. AI is increasingly providing a competitive edge in numerous sectors, including programming, content creation, and customer services. The efficiency of AI is also paving the way for a potential surge in unemployment, providing both means and context for an acceleration of this effect.

3 REDUCING MAJOR MALFUNCTIONS

The common perception of mobile phones as 'dangerous and ineffective' prevailed in their early days. Initially, their portability was a subject of ridicule, given their limited signal reception, which was largely confined to small areas within major urban centres. Despite persistent and significant concerns regarding the health risks posed by the radiation from these early models, the user base and network coverage have expanded considerably. However, it is crucial to note that two primary issues from the inception of mobile phones persist:

1. individuals sensitive to electromagnetic waves continue to seek refuge in the few remaining 'white zones', often with little societal concern for their wellbeing, and
2. cellular reception remains inconsistent across various locations.

Despite these ongoing challenges, mobile phones are widely used. They are 'good enough'.

Similarly, as of 2024, AI technology, including platforms like ChatGPT, occasionally generates erroneous responses, and it remains uncertain whether these 'hallucinations' can be entirely eliminated. AI developers know that delivering flawless technology is not a prerequisite for widespread adoption; rather, the goal is to achieve a level of functionality that meets the threshold of 'good enough' for most users. It is already 'good enough' for millions of users who feed interfaces like chatGPT with masses of personal and industrial data and provide human feedback on the results, for the technology to improve, and for capital to flow in to reduce hallucinations.

4 FEARS AND EMERGENCIES

Fear serves as a potent catalyst for social change. The use of emergency communication systems and the nomadic accessibility of crucial information often overcome entrenched hesitations. In emergency situations, the potential advantages, such as immediate access to assistance or essential information, often outweigh other considerations. This is particularly true when it concerns the safety of oneself or loved ones, especially for the most vulnerable among us.

It is evident that AI applications will be regarded as beneficial to our health and safety, akin to how smartwatches monitoring vital signs are perceived. This is especially relevant for children, the elderly, and individuals with disabilities. Notably, these populations are frequently at the forefront of technological integration trials involving body-close technologies, such as implants. These trials may pragmatically challenge deep-seated societal taboos regarding the physical integration of technologies.

5 ENHANCING HUMAN CONTACT THROUGH TECHNOLOGY

The innate human penchant for interpersonal communication has been significantly augmented by emerging technologies. Smartphones, for instance, have added a layer of immediacy to our interactions with family and friends through instant messaging and social networking. This advancement had promised enhanced human contact, facilitating communication at a rapid pace and consistency that traditional landline telephones could not sustain. The phenomenon of FOMO (Fear of Missing Out) has further amplified this trend. Among younger generations, it manifested as anxiety over missing social gatherings or updates within their social circles. For their parents, it often reflected a concern about losing connection with their children. In the era preceding the widespread adoption of the internet, marked by mass unemployment, this urgency extended to adults anxiously awaiting job opportunities via telephone calls – a scenario that preceded the ubiquity of dating apps.

Regarding AI, its role will not be to create distance from our loved ones. Instead, AI will present itself as a better method of connecting with people. AI-powered chatbots generate human-like language on demand, tirelessly and without restrictions. They have the capacity to produce language that is quintessentially human, at a volume and with a consistency unattainable by any actual human being.

The appeal of these technologies lies in their promise of increased human interaction. This very characteristic underpins their capacity to collect extensive personal data, as we frequently utilize these technologies in our most private relationships. It is evident that citizens are progressively becoming more educated about safeguarding their personal data. Some argue that the advancements in technology integration simultaneously foster social resistance towards these technologies. However, I believe this perspective confuses resistance to the implementation of technologies with the transformation of our human cultures in response to these technologies. Rather than manifesting as resistance, this cultural adaptation facilitates their harmonious integration. It fundamentally lays the groundwork for the viability of applications reliant on personal data.

At the next stage, starting with point 6, we will want to offer AI applications our most intimate and personal data. This is when the speakers of languages that are not compatible with AI will begin to be penalised.

6 CONVENIENCE AND EFFICIENCY

The statement, “I don’t like this stuff, but I dislike paperwork even more,” reflects a sentiment shared by many. A common admission is that the desire for convenience has driven the transition to mobile phones, as these devices enable the efficient handling of mundane tasks. Smartphones indeed offer an extensive array of applications and services that enhance daily life.

These include spatial navigation through GPS, time management via integrated calendars, and simplified financial transactions through mobile banking.

AI applications, designed to utilize personal data, aim to satisfy the demand for cohesive management of numerous monotonous tasks. In the healthcare sector, immediate access to personal data enables the comprehensive integration and real-time analysis of various health parameters. These include bodily metrics, dietary habits, stress levels (quantified and analysed), genetic makeup, and even environmental conditions such as the weather. This integration facilitates the generation of preliminary diagnoses, which can then be further examined by medical professionals' AI systems. The efficiency and potential benefits offered by such technological advancements are indeed noteworthy.

7 CENTRALISATION OF TOOLS

The integration of diverse services and functions into smartphones, including alarm clocks, cameras, email, messaging, social media, and even tide tables, have created the perception of liberation from a clutter of individual devices, all potentially replaceable by a single mobile phone.

AI applications, utilizing personal data, aim to personalize experiences by adapting to individual interests, behaviours, and preferences, thereby managing the plethora of apps in a manner that suits your unique style and pace. Ultimately, this seeks to liberate users from the overwhelming array of apps on their mobile devices. It seems unlikely that we will resist this trend.

8 ACCESS TO INFORMATION

The integration of the internet with smartphones has made a vast repository of information readily accessible, serving as a significant motivational factor. Individuals who were initially sceptical about mobile phones likely compared their concerns with the advantages of having immediate access to news, knowledge, and various resources.

Large corporations and governments implement information filtering mechanisms, utilizing algorithms to pre-select the information available through various channels, such as the filters currently employed by Facebook. AI, when supplied with personal data, has the power to revolutionize this filtering process. AI promises enhanced individual control, enabling users to tailor their internet experience based on trusted sources and personal interests, rather than relying on the predefined filters of others, including politicians and advertisers. Note that the upstream filtering will probably persist, but the individuals will be able to subtract further. For instance, individuals with a phobia of snakes can avoid encountering such images online. Those concerned with cultural taboos not widely acknowledged by society can navigate the internet more comfortably. Their digital assistant will exclude content on demand. Thus, each person's access to the world of information can become more personalized, shaped by their own preferences.

The next stage starts with point 9. This stage is characterised by a loss of control, points of no return and restrictions on individual choices.

9 DATA SOVEREIGNTY, SECURITY

In recent years, smartphones have evolved to include biometric authentication and advanced encryption technologies. Notably, enhancements in privacy settings, application permissions, and user control over data access have gained prominence. Initially,

privacy concerns led to hesitancy among users; however, these advancements have bolstered confidence in safeguarding personal information. Observe that this increased confidence persists despite ongoing issues, such as the persistent tracking of users' geolocation. Individuals travel daily or internationally, carrying sensitive information ranging from banking details to very intimate data in their devices, perceived as secure 'digital vaults'. Losing a smartphone is not uncommon. Certain customs authorities also legally have the right to fully access this data. This is where our situation is increasingly analogous to the paradigm of addiction, where the desire for technology overrides critical security considerations. Our security concerns are often allayed by superficial reassurances once we need access to it (or once we think we do). Consequently, we tend to neglect fundamental rules of personal security and overlook significant vulnerabilities that could impinge on our basic freedoms.

10 SOCIAL INTEGRATION

With the widespread adoption of smartphones, social pressure for their use has intensified. Friends, family, and colleagues often depend extensively on smartphones for communication and coordination. This makes it challenging for individuals to refrain from using these devices without experiencing feelings of isolation or exclusion. At this stage, technology becomes an integral part of human relationships. It not only extends these relationships by facilitating them. It is part of them, and plays a crucial role in shaping them. After such cultural mutation, human interactions deprived of technology are impaired and incomplete, lacking the support of the cultural framework that existed prior to the advent of technology.

11 TOTAL UBIQUITY OF SERVICES

A wide range of essential services, including transportation, retail shopping, ticketing, and banking, have become increasingly integrated with smartphone applications. Following the transition to digital platforms, service providers often find it costly to accommodate consumers who do not comply with these digital norms. Consequently, the access rights of these non-compliant consumers to services gradually come into question, leading to a widening of social divides. At this stage, and thinking back to the example we took with health applications, general practitioners may no longer receive training to treat patients whose personal data has not been pre-processed by AI.

12 CULTURAL AND COGNITIVE POINTS OF NO RETURN

Social practices have evolved to incorporate the use of smartphones. It is now common, and often considered polite, to confirm appointments en route, to alter the meeting location at the last minute based on preference, or to notify others of even minor delays. Such practices, integral to modern social rituals, are unfeasible without this technological aid. Generations who have witnessed this transformation recall a time when it was possible to schedule appointments months in advance and a common courtesy to arrive at the designated time and place without any need for further confirmation. Additionally, our collective ability to memorize telephone numbers has diminished as we increasingly rely on smartphones for this task. These changes represent cognitive points of no return. Organized resistance to this trend would necessitate the ability to coordinate without such tools, a skill progressively fading. The only option for individuals who resist this dependence is to reject the technology, albeit at the cost of social isolation.

CONCLUSION

This article is not a prophecy but rather a prospective analysis. In this narrative of our integration into the mobile technology revolution, is there truly a fundamental difference, either contextually or technologically, that distinguishes it from AI? From my perspective at the beginning of 2024, I believe we have moved largely beyond phase 3. Major industries are evidently

preparing for subsequent developments, and I perceive no barriers to this ongoing evolution of technology integration, reminiscent of this other technological revolution I have witnessed in my adult life. If, like me, you discern no fundamental disparity between the societal adoption of these two technologies—mobile phones and AI—it seems logical to infer that resistance to AI is unlikely. The influence of this wave will rapidly extend to our most personal spheres, likely impacting even our physical bodies and cognitive processes. Any individual or collective resistance will amount to mere cultural adaptations to a new reality. This evolution presents and will continue to present multiple challenges, which the scientific community is currently ill-equipped to address [1]. AI technologies demand substantial hardware capabilities for processing vast amounts of data. A limited number of competing companies are making significant decisions that affect our lives and our future as a species. These decisions are made without democratic input and without the necessary internal resources to fully consider the wide-ranging societal implications of their actions. These companies do not employ scientists from the humanities to begin to address these problems.

As a linguist, it is the danger of a drastic reduction in linguistic diversity that appears particularly salient to me. To illustrate with a concrete case, the French state has a single official language, French, which will undoubtedly be incorporated into AI technologies. The more than one hundred other unofficial, unprotected languages of the citizens of the French state are not at all equipped to overcome the AI barrier. Most of them are in a state of digital insecurity, which we are still struggling to map [2].

There is a widespread belief in our societies, apart from within a few influential but isolated circles, that resistance to artificial intelligence will be strong in society, and that forecasts of technological revolution have erred by not taking sufficient account of the resistance it will trigger. In Brittany, I come across this belief in very different generations and very different social circles, from the Monts d'Arrée to the high-tech academic circles of the megacities, as well as in... the high-tech circles of the Monts d'Arrée. I respect this belief as such. As a belief. I respect the emotions it protects, and the people who feel them. But I think it is objectively inaccurate. I also think that this belief is dangerous when it slows down the measures we can still take to adapt to this coming mutation. This change is objectively inevitable. The speakers need to be aware of the particular dangers that the development of AI will pose to their linguistic practices if the new tools cannot work in their languages. Denying it slows down the implementation of emergency measures to counter these dangers. If our languages cannot be processed automatically, so as to integrate the new tools, they will face an even more radical reduction in practice than they did in the twentieth century. Most of them will disappear.

Anyone who knows anything about the history of minority languages in the twentieth century should be on high alert right now. Here's why. We know that languages disappear when parents think that these languages are dangerous for their children. Non-AI compatible languages will become dangerous for their users, as the integration of the new apps progress in society. We have seen that there will be AI applications to which we - ourselves - will want to offer our intimate and personal data. These are points 6 to 8 illustrated above. If small languages cause these applications to malfunction, we will replace them with the languages that let them function properly. This is already the case in practice when we simply dictate an SMS in French to take advantage of its automatic writing rather than typing it in Breton. But the scale of avoidance is going to be immense, unparalleled. When society reaches the points of no return, the points between 9 and 12 of total and irreversible integration, the pressure to eliminate malfunction sources will increase. Nonconformity will come with social exclusion, because these tools settle at the heart of human relationships. When these applications become essential for access to banking, justice, education, dating and health services, we will seek to protect our children's future by reluctantly forbidding them to use our non-AI compatible

languages, which will be objectively transformed into a serious social handicap, in a great global movement towards monolingualism. History will say that we consented to this.

There are still between 6,000 and 7,000 languages in the world. They are the products of our human bodies, of billions of human brains interacting with each other over thousands of generations. Each disappearance of a language produces deep wounds, and echoes over several generations. Imagine being raised by people who refuse to speak their own native language to you. And we are very bad at healing those wounds. Our generations have the opportunity to mobilise to prevent this harm to be done, to ensure that this diversity is not wiped out, and that AI is an opportunity for multilingualism and the preservation of this incredible diversity. It undoubtedly has the power to be for languages made compatible with this technology.

We need a global rescue plan, a lively public, scientific and societal mobilisation, a collective rescue plan for our linguistic treasures. In practice, languages need a digital data set of written and spoken language, put into the right form for digital processing [3]. It is now crucial and urgent for our language communities to build up corpora of written and spoken language and make them accessible and open, to scientists and industry alike, with open copyright. You might think that the urgency derives from their actions in the first place, and you might also be very angry. However, preventing them from accessing the AI integration material for your languages is resistance of a purely moral nature, and is not effective in improving the situation. Resistance to AI is not going to happen. Taking languages hostage would seal their fate. Swimmers know that when the wave is too big, you have to dive in. Not just stand there and say that the wave shouldn't be there. Language support policies need to rethink their arsenal of support by organising them around the necessary surviving digital toolkit [4]. They need to encourage the emergence of trans-disciplinary collaboration cells between people who know how to code, to educate AI in languages with restricted corpora, and people who know how to talk to people, harvest language material in an ethical manner, and enrich it with precise standard annotations. Language communities need to be digitally empowered, in order to build the digital resources they need to be integrated into the new tools.

As for the scientific communities that produce, process and work on corpora, my very dear colleagues, descriptivist and formal linguists, field workers, corpus and elicitation linguists, annotators and engineers of data architecture, we have an historic opportunity to safeguard the object of our research for human societies.

It's time to rise to the occasion and show them what we can do.

ACKNOWLEDGEMENTS

I would like to thank Rayan Ziane, Loic Grobol, Reun Bideault and Milan Rezac for their critical reviews and discussions on the matter. Any misinterpretations and shortcomings remain my own.

REFERENCES

- [1] Samuel R Bowman. 2023. Eight Things to Know about Large Language Models. Apr. 02, 2023. arXiv:2304.00612.
- [2] Mélanie Joutiteau, Sylvain Kahane and Loic Grobol. 2023-present. Entrelangues, second edition, IKER & Modyco, CNRS. Retrieved from <https://entrelangues.modyco.fr/>
- [3] Steven Abney and Steven Bird. 2010. The Human Language Project: Building a Universal Corpus of the World's Languages. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics ACL, Uppsala, Sweden, 88–97.
- [4] Mélanie Joutiteau. 2023. Guide de survie des langues minorisées à l'heure de l'intelligence artificielle : Appel aux communautés parlantes. Lapurdum, numéro spécial 6. Retrieved from <https://hal.science/hal-04090195v2>