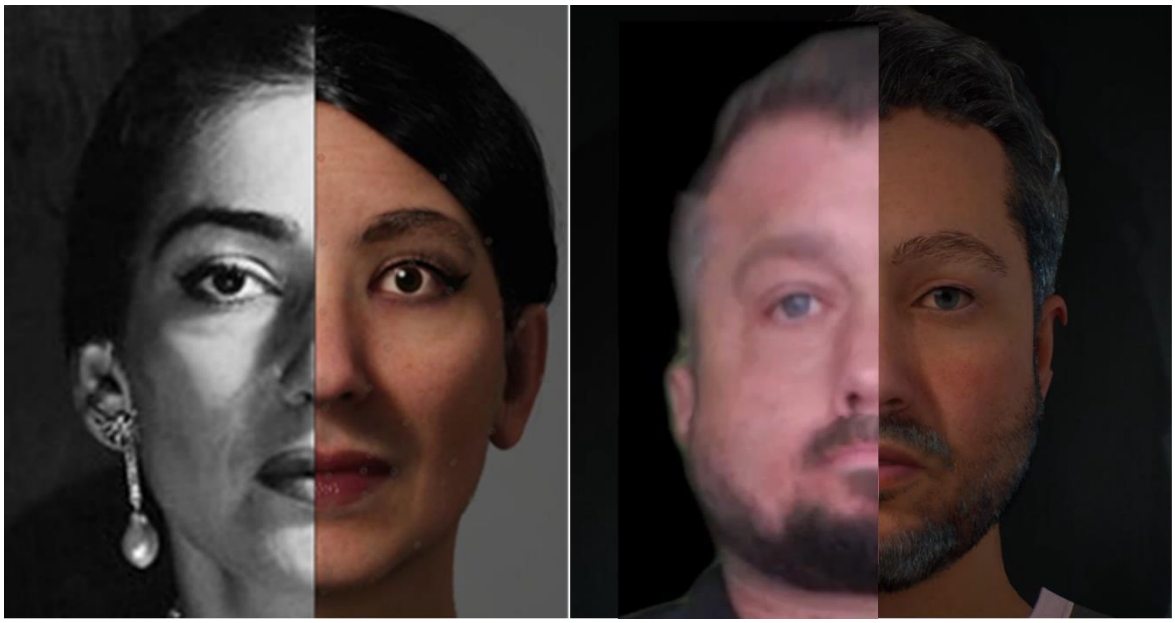
Synthetic Humans
Industry Technical Review 2023

Michael Davey[1], Roberto Iacoviello[2], Mark Applin[3], Martin Monov[4],
Dennis Laupman[5], Rustem Vilenkin[6]

[1] Independent Technologist, [2] Rai, [3] Signly, [4] White Light (a D&B Solutions company),
[5] Pluxbox, [6] Respeecher



Contents

Contents 2
About this report 5
Acknowledgements 5
Executive Summary – Entertainment 7
Executive Summary - Sign Language 7
Executive Summary - Lip-readability 7
Overview - Entertainment 8
Challenge & Innovation 9
Overview - Accessibility 10
Sign Languages 12
Dialects 12
Translation and Interpretation 13
Interpreters 15
Translators 16
State of the Industry 17
Legal framework 17
Broadcast, media and entertainment industry 18
Solution Paradigms 20
Generative AI Video 20
Virtual Actors & Virtual Production 20
Motion capture technology 21
Special media-friendly signs 22
Translation of British Sign Language 22
Our approach 22
Key challenges 22
IBC Lip Sync Battle: Bridging the Gap with Sign Language Innovation 24
Impact 24
Conclusions 25

Key learnings and future work....27
Creating avatars from archive audio and video footage....27
Motion capture and cleanup....27
Maturity of ecosystem....28
Speech-to-speech synthesis....28
Media supply chain automation....29
Avatar modelling and sculpturing....30
Lip sync....30
Lack of broadcast standards....31
Focus groups....32
Commercial frameworks....32
The evolving legal and commercial landscape....33
Identification and remuneration....33
Summary....33
Appendix A – Policy Statements....34
Environmental Statement....35
AI Statement....37
Appendix B – Platform Descriptions....40
High Level Technology Overview (BT):....41
Version 0.2 MR (BT)....41
Revision History....41
Introduction....41
Background Brief....41
Overall Approach....41
Contribution Approach....42
Distribution Approach....42
Presentation Approach....42

About this report

This report covers the technical challenges involved in Synthetic Humans for Entertainment and Accessibility.

The goal is to use common tools to create super-realistic digital copies of humans for TV and other platforms. This could shake up how entertainment and accessibility are delivered.

We considered two use-cases:

1. Entertainment: Make a lifelike Maria Callas in both appearance and voice.

2. Accessibility: Focus on voice synthesis and lip-sync to make content more accessible for the growing amount of content.

Across these two use-cases, we identified three key goals:

- Making 3D models that look real.

- Getting these models to interact seamlessly with existing footage.

- Incorporating everything into TV production workflows.

The challenges here include voice synthesis, lip-sync for lip-reading, and even British Sign Language weather forecasts.

Acknowledgements

The authors wish to acknowledge everyone who has talked with us or written about their experiences with the subject matter for this paper; we have extensively drawn on their experiences and works to produce this paper.

We have benefitted from comments and critique on early versions of this paper from a number of readers including Ian Wagdin (BBC), Lisa Connelly and Will Kreth from HAND (Human and Digital), Roch Nakajima (Noitom), ErinRose Widner (Verizon), David Reilly, Jouni Frilander and Petri Karlsson (YLE), Gregg Young (VRT), Lisa and Natasja (4DR Studios), Paola Sunna & Bram Tullemans (EBU), B.K. Johannessen (Epic Games/Unreal), Ron Martin (Unity),

Joao Felix (V-Nova), Oona Patterson (NDCS), Cate Calder (CSUK) and Colin Christie (Acoustic Reality/University of Southampton).

Special thanks to Matt Kirby (Deaf actor and sign language translator) for being our source talent for our Accessibility avatar (Matteus); to Muki Kalhan, and Mark Smith (IBC) for their mentoring; and to Abigail (Aby) Kingsland and the IBC Accelerator team for their support.

Executive Summary – Entertainment

Over the past few years, the field of synthetic humans and digital twins has experienced a significant rise in research and production. These innovative concepts hold the promise of fundamentally transforming our daily lives and professional endeavours, generating substantial attention and discussion among both experts and the public alike. As of 2022, the worldwide market for digital human use-cases had reached a substantial value of 29.51 billion USD[1].

Executive Summary - Sign Language

Sign language is the native language of the Deaf community who embrace it as part of their culture. Yet most information and communications are never translated into sign languages, making it difficult for deaf people to access. The chronic scarcity of sign language translators and interpreters, the rising demand for translations, and the need for scalable solutions means that needs can only be met through machine-generation of sign language translation video.

Significant advancements have been achieved in developing technology solutions, especially in the realm of Artificial Intelligence. However, there are still some scientific and engineering challenges to overcome, specifically regarding motion capture and clean-up for sign language performances and fitting translations into the allocated time frame of the source.

Executive Summary - Lip-readability

Lip-readability refers to how easily someone can understand spoken language by watching the speaker's mouth movements, facial expressions, and gestures. It's particularly important for people who are hard of hearing or deaf. Good lip-readability means clear and anatomically nuanced mouth and face movements, proper lighting, and straightforward language to help interpret what's being said.

[1] Emergen Research, July 2023: https://www.emergenresearch.com/industry-report/digital-human-avatar-market

Increasing lip-readability can make content more accessible, and that fits right into the whole idea of inclusive design.

Overview - Entertainment

Digital humans provide content creators with unparalleled creative flexibility and control. Filmmakers and animators can manipulate and fine-tune the appearance, behaviour and narrative arcs of avatars to fit their artistic vision. This versatility allows for exploration of diverse storytelling techniques, character developments, and imaginative worlds that can be otherwise challenging or costly to achieve with conventional methods.

The growing demand for realistic, high-quality photorealistic characters for high-budget film productions has created the need to develop efficient production techniques that can create these products quickly and cost-effectively.

The construction of 3D digital models of human beings from sensory data has long been an issue for computer vision and graphics. Although much work has been done on the development of various acquisition hardware and reconstruction algorithms, traditional reconstruction pipelines are still characterised by expensive acquisition systems and cumbersome acquisition procedures that make them difficult to access. In addition, hand-crafted pipelines are prone to reconstruction artefacts, which reduce their visual quality.

Even though broadcasters want to use new technologies and creative solutions to improve production quality, they often find that these solutions don't easily fit into a traditional broadcast pipeline's timeframe, budget, resources or skills. Furthermore, there is a lack of knowledge on the best practices for integrating digital humans into broadcasted television productions.

There is therefore a need to search for new methods to create realistic synthetic human beings that can be used in broadcast television productions at an affordable price while meeting quality standards.

Challenge & Innovation

This project is driven by the ambitious goal of crafting exceptionally lifelike synthetic humans of the highest quality. To achieve this, we are actively seeking out innovative and cost-effective workflows that can operate within the confines of rigorous time and budget constraints.

Furthermore, we aspire to exploit the broadcaster's archive to uncover and bring to light stories that have not been previously considered noteworthy.

Moreover, we want to deploy these synthetic humans across a wide spectrum of platforms and devices. This includes not only conventional websites but also emerging frontiers like the volumetric ledwalls and head mounted displays. The digital landscape has evolved significantly, and we recognize the need to adapt. Therefore, we are keen to integrate these synthetic entities seamlessly into the realms of multimedia, extended reality, and mixed reality experiences.

In summary, our project is a comprehensive endeavour, blending cutting-edge technology with creative storytelling, all while working within the constraints of time and budget. Our objective is to usher in a new era where synthetic humans assume a critical role in providing captivating narratives across a variety of digital platforms.

In our pursuit of creating an accurate digital twin, we've identified seven key challenges and areas for investigation, each requiring innovative solutions:

1. Data Collection & Management: We embarked on extensive research, meticulously gathering information, photographs, video recordings, and audio clips from broadcaster's archives. Leveraging internal search and retrieval software, we ensured a robust foundation for our project. Managing diverse technologies and standards across our workflow posed a significant integration challenge. Ensuring data quality, consistency, and reliability was another hurdle, addressed through techniques like super-resolution algorithms to enhance old archive videos.

2. Facial Resemblance: To achieve an authentic likeness, we employed a multi-step approach. After selecting 2D pictures, we utilized Facebuilder to generate the face mesh, subsequently refining it in Metahuman and enhancing details with Substance Painter. While automation played a role in creating a basic mesh, human intervention was essential to achieve the

desired level of accuracy. We explored model validation methods, including heat maps, to assess differences between 2D pictures and 3D models.

3. Body Movement: Achieving convincing body movements was a multi-pronged effort. While fully automated workflows from raw motion suit and live puppeteering motion data provided success in many areas, a number of specific challenges including resolving arm-body interpenetration and realistic simulation of clothing movements required time-consuming manual cleanup and animation techniques.

4.Lip Sync: We encountered difficulties with AI-based software for lip sync and facial expressions, primarily designed for speech rather than singing. As a result, we opted for a real performer and live acquisition, benefiting from the maturity of voice cloning technology.

5.Audio Enhancement: To tackle audio challenges, particularly in reverberation, we invested in learning and refining our audio processing techniques.

6. Automation of the Process: Recognizing the resource-intensive nature of our project, we focused on automation. We developed a hybrid orchestrator to guide 3D artists through the complex journey from 2D images to 3D models. This tool not only streamlines the process but also assists employees who may be new to the technology or resistant to change.

7. Multi-platform Use: We harnessed the potential of our 3D model across different platforms, including video output, volumetric ledwall and head-mounted displays. However, interoperability issues arose due to the lack of standards, emphasizing the importance of adaptability and reusability of our 3D model.

In the face of these challenges, our project demonstrates the power of innovation, collaboration, and adaptability in creating a high-fidelity digital twin that captures the essence of Maria Callas. We look forward to further advancing our capabilities and overcoming hurdles as we continue to push the boundaries of technology and storytelling.

Overview - Accessibility

The chronic scarcity of sign language translation professionals, and an accelerating demand for translations dictate an innovative solution to meet

demand. Conventional machine learning approaches are hampered as the range, volume and quality of data, and the complexity of data preparation, make the application of Generative Adversarial Networks (GANs) unsuitable for production-grade work, which is not necessarily obvious from their alluring demonstrative capability.

The state of the art is to apply avatar technology, applying motion-captured translations, assembled by bespoke Artificial Intelligence (AI), into a multi-layered performance that covers not just the signs, but also non-manual features, expressions, and emotion. This is the only approach at present that offers a feasible roadmap to deliver accurate, high quality sign language videos at scale.

Translations of video and audio content, unlike text-based content, present additional challenges relating to timing: sign languages typically offer a different rhythm and cadence to spoken languages, making direct translations extremely difficult to match to the timing of the source. Pre- and post-production considerations offer varying degrees of resolution.

Broadcasters face several challenges to make content lip-readable:

- Camera Angles: A bad angle can obscure the mouth, making it tough to lip-read.
- Lighting: Poor lighting can create shadows that interfere with visibility of facial movements.
- Fast Dialogue: Rapid speech makes it hard to keep up with lip-reading.
- Multiple Speakers: Quick switches between speakers can disrupt focus.
- Visual Obstructions: Logos or tickers can block the view of a speaker's mouth.
- Audio Sync: If the audio and video aren't perfectly synced, it confuses aural lip-readers.
- Dialect and Accent: Variations in speech can make lip-reading tricky for some viewers.
- Technical Limitations: Real-time broadcasts may not always allow for adjustments that improve lip-readability.

Sign Languages

There are more than 200 sign languages in common use throughout the world[2], bringing their own syntax and grammar distinct from the spoken languages of their home nations. In some cases, there is regional variation in a sign language, and sometimes multiple sign languages in use. For instance, British Sign Language and American Sign Language are different, and the sign language used in Quebec, Canada (Quebec Sign Language) is different to American Sign Language, British Sign Language and French Sign Language.

In the United Kingdom, British Sign Language was officially recognised by the UK Government in 2003[3] and this recognition was protected in law in 2022 with the passing of the 2022 BSL Act[4]. Many other countries have independently passed similar laws.

Many people who are born without hearing, and many who are deafened in infancy, struggle with literacy as learning to read typically requires the support of phonics[5,6]. For this reason, captions and subtitles are not an effective or accessible means of communication for many d/Deaf people.

Dialects

Sign languages have regional dialects, analogous to regional dialects in spoken languages. These dialects are influenced by factors such as geography, culture, and historic events. Signs used in BSL in Scotland, the North of England, East Anglia, Wales, London and the Southeast may all be different.

---

[2] Johnston, T. (2019), 'Sign languages of the world: A comparative handbook'; Walter de Gruyter GmbH & Co KG.

[3] UK Government statement (cited in Hansard) (2003), 'Written Statement on British Sign Language by Mr Andrew Smith MP, Secretary of State for Work and Pensions', *https://hansard.parliament.uk/Commons/2003-03-18/debates/52ac28bc-48db-4cf3-a6e5-6415f94cd7ae/BritishSignLanguage*

[4] UK Government (2022), 'British Sign Language Act 2022', *https://www.legislation.gov.uk/ukpga/2022/34/contents/enacted*

[5] Marschark, M., & Hauser, P. C. (2012), 'Deaf cognition: Foundations and outcomes'; Oxford University Press

[6] National Deaf Children's Society (2020), 'Literacy and learning for deaf children', *https://www.ndcs.org.uk/information-and-support/literacy-and-learning/*

While there may be regional differences in sign languages, the fundamental grammar and syntax of British Sign Language mostly remain the same. As with spoken languages, sign languages are complex and sophisticated systems of communication that evolve and adapt to their social and cultural contexts.

Translation and Interpretation

In general, interpretation deals with spoken language in real time while translation focuses on written content. Sign language translation and interpretation services companies rely on human interpreters and translators. They are constrained by a worldwide shortage of sign language translators and interpreters[7,8,9,10], resulting in significant prioritisation of resource allocation according to perceived importance[11,12,13,14]. Recruitment and training of new translation service personnel is impeded by the perceived and actual difficulty

---

[7] World Federation of the Deaf (2019), 'Sign Language Rights for All', *https://wfdeaf.org/wp-content/uploads/2019/09/WFD-SLR4All.pdf*

[8] International Federation of Translators (2013), 'The Global Demand for Sign Language Interpreting Services', *https://www.fit-ift.org/wp-content/uploads/2013/11/FIT-FL-Report-2013_EN.pdf*

[9] American Sign Language Teachers Association (2018), 'Interpreter Shortage Crisis', *https://www.aslta.org/interpreter-shortage-crisis/*

[10] World Health Organization (2018), 'Improving Access to Health Services for Persons with Disabilities', *https://www.who.int/news-room/fact-sheets/detail/disability-and-health*

[11] Registry of Interpreters for the Deaf (2017), 'Code of Professional Conduct', *https://www.rid.org/ethics/code-of-professional-conduct/*

[12] National Association of the Deaf (2015), 'Guidelines for Effective Communication with Deaf and Hard of Hearing People', *https://www.nad.org/resources/american-sign-language/community-and-culture-frequently-asked-questions/guidelines-for-effective-communication-with-deaf-and-hard-of-hearing-people/*

[13] Gallaudet University (n.d.), 'What is a Sign Language Interpreter?', *https://www.gallaudet.edu/sign-language-interpreting/what-is-a-sign-language-interpreter*

[14] Sign Language Interpreters and Transliterators Association (2019), 'Position Paper on Prioritization of Requests for Interpreting Services' *https://sliata.org/wp-content/uploads/2019/07/Prioritization-of-Requests-for-Interpreting-Services-Position-Paper-SLIATA.pdf*

and costs incurred, exceeding the duration and costs of an undergraduate degree[15,16,17,18,19,20,21,22,23].

British Sign Language is a complete, distinct language, independent of English. It differs not just in using visual rather than audible vocabulary, but also in grammar and syntax, and in the very rules of articulation. A convincing and compelling translation requires more than simply identifying the correct signed words and putting them in the correct sequence.

[15] International Federation of Translators (2013), 'The Global Demand for Sign Language Interpreting Services', *https://www.fit-ift.org/wp-content/uploads/2013/11/FIT-FL-Report-2013_EN.pdf*

[16] National Interpreter Education Center (n.d.), 'Cost of Training', *https://www.interpretereducation.org/cost-of-training/*

[17] Chartered Institute of Linguists (2015), 'Language Services Market Survey', *https://www.ciol.org.uk/sites/default/files/Language-Services-Market-Survey-2015-Report.pdf*

[18] National Register of Public Service Interpreters (2017), 'Recruitment and Retention of Public Service Interpreters in the UK', *https://www.nrpsi.org.uk/wp-content/uploads/2017/06/NRPSI-Report-Recruitment-and-Retention-of-Public-Service-Interpreters-in-the-UK-2017.pdf*

[19] UK Government (2019), 'National Framework Agreement for the Provision of Interpreting, Translation and Transcription Services', *https://www.gov.uk/government/publications/national-framework-agreement-for-the-provision-of-interpreting-translation-and-transcription-services*

[20] Signature. (n.d.), 'Become a Sign Language Interpreter', *https://www.signature.org.uk/become-a-sign-language-interpreter/*

[21] Association of Sign Language Interpreters (n.d.), 'Training to Become an Interpreter', *https://www.asli.org.uk/training-to-become-an-interpreter*

[22] National Registers of Communication Professionals (n.d.), 'How to Register', *https://www.nrcpd.org.uk/How-to-Register*

[23] Heriot-Watt University (n.d.) 'Interpreting: Introduction', *https://www.hw.ac.uk/schools/social-sciences/departments/languages-intercultural-studies/study/interpreting.htm*

Interpreters

In the United Kingdom, there are more than 150,000 British Sign Language users[24,25,26], and only around 1,000 Registered Sign Language Interpreters (RSLI)[27,28,29], each of whom can translate for only 4 hours per 8-hour working day due to the cognitively exhausting nature of the work[30].

---

[24] Welsh Deaf Council figures cited in UK Government Department for Work and Pensions (2017), 'Market review of British Sign Language and communications provision for people who are deaf or have hearing loss', *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630960/government-response-market-review-of-bsl-and-communications-provision-for-people-who-are-deaf-or-have-hearing-loss.pdf*

[25] BDA figures cites in Northern Ireland Assembly (2020) paper, "Sign Language Legislation", http://www.niassembly.gov.uk/globalassets/documents/raise/publications/2017-2022/2020/communities/7720.pdf

[26] British Deaf Association (n.d.), 'BSL Statistics', *https://bda.org.uk/help-resources/#statistics*

[27] The British Deaf Association (n.d.), 'How many qualified BSL Interpreters / Translators are there in the UK?' cites Signature (2015) which cites NRCPD: "there are 908 registered sign language interpreters (RSLI) and a further 234 trainee sign language interpreters (TSLI) in the UK. There are 11 registered sign language translators.", https://bda.org.uk/help-resources/#statistics

[28] National Registers of Communications Professionals working with Deaf and Deafblind People (NRCPD), referenced in UK Government – Department for Work and Pensions (2017), 'Market review of British Sign Language and communications provision for people who are deaf or have hearing loss', https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630960/government-response-market-review-of-bsl-and-communications-provision-for-people-who-are-deaf-or-have-hearing-loss.pdf

[29] National Registers of Communications Professionals working with Deaf and Deafblind People (NRCPD), 'Registration figures', https://www.nrcpd.org.uk/registration-figures

[30] Association of Sign Language Interpreters (2023), 'Working with an Interpreter and a Translator – Booking an Interpreter or Translator', *https://asli.org.uk/working-with-an-interpreter/*

Translators

NRCPD (2023)[31] reports that there are only 33 registered Deaf translators (RSLT) nation-wide. The acceleration of the rate at which non-signed content is created compounds the problem. For example, last year the BBC released 28,000 hours of new content[32]. Every single hour, tens of thousands of new web pages are crafted[33] and 30,000 hours of new videos are uploaded to YouTube[34].

Machine translation is essential for sign language users to gain equality of access to information and entertainment. Such solutions are classified as deep technology (*deep tech*) as they are based on substantial scientific and engineering challenges.

---

[31] National Registers of Communications Professionals working with Deaf and Deafblind People (NRCPD), 'Registration figures', https://www.nrcpd.org.uk/registration-figures

[32] According to the BBC's Annual Report, they broadcast 130,000 hours of television content per year and 28,000 hours was new, original TV content (21.5%). BBC (2022), 'BBC Group Annual Report and Accounts 2021/22' and 'BBC Commissioning Supply Report 2021/22', *https://downloads.bbc.co.uk/aboutthebbc/reports/annualreport/ara-2021-22.pdf* and *https://downloads.bbc.co.uk/commissioning/site/bbc-commissioning-supply-report-202122.pdf*

[33] While it is difficult to determine the exact number of new web pages created every hour, the site Worldometers provides an estimate. Based on their data, an estimated 547,200 new websites are created every day, or approximately 22,800 new websites per hour. Source: Worldometers (2021), 'Internet Users by Country (2021)', https://www.worldometers.info/internet/

[34] Statista (2023), 'Hours of video uploaded to YouTube every minute as of February 2022', *https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute*

State of the Industry

Established sign language translation companies continue to struggle with single source motion capture and scalability issues.

In the last 4 years, several AI Sign Language companies have emerged. Notable is Huawei's demonstration of automatic translation from Mandarin and Cantonese into Chinese Sign Language (CSL, ZGS). Several ASL startups have appeared in the US, including signall.us, slait.ai and gosign.ai. Closer to home, humanid.dk (Denmark) are reported to be working on an AI sign language translator for Danish Sign Language (Dansk Tegnsprog or DTS).

In the last 4 years, Epic Games has announced MetaHumans[35] and Unity has purchased Ziva Dynamics[36]. Microsoft, Meta (Facebook), NVIDIA, Soul Machines, Digital Domain, Pinscreen, Genies, Amazon, Google and others are all working in the hyper-realistic avatars space – competing for superiority, arguably with Soul Machines and Unity in the lead at the time of writing.

On the standardisation front, there is interesting work currently being undertaken for Object-Based Media (the OBM Discussion Group[37] and the DVB OBM Working Group[38]) and for the Metaverse (Metaverse Standards Forum). This work is relevant to the right-hand-side of our pipeline, particularly the renderer. We've found that by adopting Pixar's USD as our native 3D scene descriptor, we can leverage a wide range of creative tools from many different suppliers.

There has been significant progress with developing algorithms and systems for natural language processing, image and video processing, speech recognition, speech synthesis and so on. Progress with Transformers (e.g., GPT, BARD, LLaMA) is particularly impressive, providing an expensive way to summarise or condense text to any arbitrary length desired.

Legal framework

There is an evolving legal framework and sustainability, involving government discussions, consultations and proposed legislation. This project has created a reference AI statement and Environmental statement based on industry best

---

[35] https://www.unrealengine.com/en-US/metahuman

[36] https://www.youtube.com/watch?v=xeBpp3GcScM

[37] https://www.linkedin.com/groups/9107257/

[38] https://dvb.org/news/dvb-seeks-input-on-object-based-media/

practices. These policy templates are free for the industry to use if they wish and are available in Appendix A.

2022 saw an update to the Health and Social Care Act[39] and the passing of the British Sign Language Act 2022. The latter provides a framework for the use of BSL in public spaces and mandates that public announcements be accessible in BSL.

2023 saw several unions striking for reasons related to royalties and residuals payments, particularly when AI has been involved in creating or editing a work or performance. Strikes have been particularly notable in the USA, where the Writers Guild of America (WGA) and the Screen Actors Guild and American Federation of Television and Radio Artists (SAG-AFTRA) have been striking.

2023/2024 is expected to see the passing of an update to the UK Digital Economy Act 2017, introducing a minimum level of BSL content availability for on-demand services (implementing clause 93 of the aforementioned bill).

In the same timeframe, both the UK Government[40] and the EU[41] are progressing with their respective AI regulations.

Broadcast, media and entertainment industry

The broadcast, media, and entertainment industry has been going through significant changes and challenges due to the COVID-19 pandemic, global recession and the impacts of Brexit. However, the industry has shown resilience and adaptability in response to these challenges.

Streaming services have continued to grow in popularity, with major players such as Netflix, Amazon Prime Video, Disney+, HBO Max, BT, Sky, and Liberty Global (Virgin Media O2) competing for subscribers. The pandemic has accelerated the trend of cord-cutting and shifted consumer viewing habits towards streaming services.

Traditional broadcast television networks and cable providers have faced declining ratings and revenue as viewers increasingly turn to streaming platforms. This has led to consolidation and mergers among media

[39] https://www.legislation.gov.uk/ukpga/2022/31/contents/enacted
[40] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach
[41] https://artificialintelligenceact.eu/

companies, such as the recent merger of Discovery and WarnerMedia and subsequent joint venture with BT Sport.

The Coronavirus pandemic catalysed the entertainment industry's adoption of new ways of producing and distributing content. Remote production and virtual events have become more prevalent, and there has been a greater emphasis on creating content for online platforms and social media.

The rise of social media influencers and user-generated content has disrupted the traditional model of celebrity endorsement and advertising. Brands are increasingly turning to influencer marketing and collaborations with creators to reach younger audiences.

The industry is also grappling with issues of diversity and representation, with movements such as #MeToo and Black Lives Matter leading to greater scrutiny of hiring practices and content production. Notably, media, broadcast and entertainment companies are well represented as Valuable500 members (BBC, Bloomberg, BT, Canal+, Channel 4, ITV, Sky, LibertyGlobal and WarnerDiscovery amongst others).

Overall, the broadcast, media, and entertainment industries are in a state of flux, with traditional players being forced to adapt to changing consumer preferences and new technologies. The pandemic has accelerated many of these trends, and it remains to be seen how the industry will continue to evolve in the coming years.

Solution Paradigms

There are two main approaches to solving this machine translation problem: *generative AI* and *virtual actors*. We have taken the latter approach for reasons that this document explores.

Generative AI Video

Generative AI is a subfield of artificial intelligence that involves creating AI systems that can generate new, original content. Generative AI models work best when the output required is very similar to the training data available. They are trained on large datasets of existing content and learn to generate new content by identifying patterns and relationships within the data.

These models use techniques such as neural networks, deep learning, and probabilistic programming to create new content that may be near-indistinguishable from content created by filming humans.

Generative AI presents several challenges when used for sign language translation. These challenges include the need for large datasets which can be difficult to obtain, expensive scaling of the AI, ethical privacy, bias and trust concerns, legal issues, quality control, difficulty of validation, and inflexibility of the AI.

Despite the challenges, there have been efforts to address the data scarcity problem and create a comprehensive, open corpus of sign language signs.

However, creating a Generative AI that can translate a variety of source material into sign language robustly, accurately, and consistently is an extremely complex task.

Virtual Actors & Virtual Production

Virtual actors (and more generally, virtual production) are not subject to the problems described above. In addition, they afford a large degree of artistic freedom in terms of the final product in terms of lighting, camera angles, camera movements, composition, editing, colour, staging, aesthetics of the actors themselves and their attire, etc. Creatives can very quickly experiment with different visual styles and quickly make changes to scenes in real-time.

However, they also present their own set of challenges such as the uncanny valley, lack of authenticity, limited range of expressions, and dependence on proprietary technology.

Recent technological advances have made creating high-quality virtual actors both cost and time effective, and it is expected that third-party avatars will match or exceed the realism of Deep Fakes by 2023 or 2024.

Motion capture technology

Motion capture technology has been used in the film industry since the 1970s and its cost has decreased since then, allowing for its use in various industries including video games, sports, medicine, and virtual reality. The price of motion capture equipment ranges from a few thousand to millions, depending on the size of the studio and its purpose.

Currently, there are no ready-made motion capture solutions designed specifically for sign language.

Motion capture solutions can be divided into five main categories: optical marker motion capture, inertial motion capture, volumetric RGBD, volumetric RGB and rotoscoping.

1. Optical marker motion capture uses multiple cameras to track reflective markers placed on a subject's body.
2. Inertial motion capture uses sensors attached to the body to measure orientation and movement.
3. Volumetric RGBD uses computer vision techniques to process images from RGBD cameras, which record colour and depth information.
4. Volumetric RGB uses standard video cameras to capture volumetric movement, often with the help of AI techniques.
5. Finally, rotoscoping is an animation technique that uses tracing paper or digital packages to trace keyframes, with the software interpolating for the in-between frames.

There are two emerging machine learning-based volumetric motion capture systems - move.ai and Plask - which use multiple RGB+LIDAR cameras and AI to improve accuracy and efficiency.

Reports from the video games industry suggest move.ai is superior to low-cost inertial solutions, as it does not require gloves, has better accuracy, requires less clean-up and has an intuitive user interface.

While move.ai is designed to handle occlusion, its ability to deal with complex hand/finger interactions and signing nuances is yet to be tested by us. Several

research institutes and companies - including Facebook Research, Microsoft Research, Intel Studios and XSens - are actively researching this area.

Special media-friendly signs

Media-trained individuals use special signs that are a combination of regional dialects, creating a "Received Pronunciation" of British Sign Language (BSL). These compound signs are not generally used by Deaf people but are understood by most. Examples include the BSL sign for 'from', which is either the dominant hand, held palm up in 'B' shape, with fingers pointing away, or the dominant hand, held palm down in 'B' shape, above the non-dominant hand, held with palm facing right in 'B' shape. Media-trained interpreters will often adopt a compound sign which is a combination of these two.

Translation of British Sign Language

Translation of British Sign Language is a difficult task due to the complexity of the language and its differences from English. Direct word-for-word translation is not possible, and meaning and context can be altered or replaced by facial expressions, head movements and arm movements. Translation of speech from video to BSL is constrained by timing, as the articulation must fit within the overall time available, without distorting the meaning. Quoting someone also poses a challenge, as the translation must fit into a fixed time without compromising the flow of content.

Our approach

Our approach translates source language to destination language using a bespoke language model. A time-series notation produced from this model is then processed by the sequencer in the avatar engine to pick and place the correct hand animations, body and facial expressions, and lip patterns onto the timeline. Motion capture techniques are used to create a wide range of movements, from naturalistic to highly stylized, and can be customized and reused to suit the needs of a particular project. The pipeline is designed for this generated performance, by adding morphemes to the translation corpus and the vocabulary corpus and cleaning, transforming, labelling and inserting the animation data and its associated metadata into the corpora.

Key challenges

Today, the key challenges remaining are:

a) to improve the efficiency of performing and cleaning up motion capture so that it is quicker and easier to acquire new sign language vocabulary
b) to improve our ability to fit a translation into the available time
c) to develop industry-specific standards and regulations, including the creation of digital twins that are universal across tools, technologies and use-cases
d) to clarify issues of content ownership and licensing when using digital twins in live-stream and broadcast scenarios to prevent legal complications and disputes

There are also several areas where improved workflow performance and automation would drive adoption of the technology. Improvements to the Volumetric Capture pipeline could resolve the data acquisition, synchronisation and fragmentation issues and significantly reduce the amount of manual clean-up for facial captures and hand movements. This is particularly important when considering live streaming and broadcasting needs/productions.

Similarly, implementing appropriate network infrastructure, high-performance computing and other scalability considerations to ensure that low latency and performance is achieved, is also particularly important for live streaming and broadcasting needs/productions.

In addition, incremental advances shall continue to improve the quality and authenticity of translations, to the realism, emotions, expressions and use of body language, and continue to develop the scalability of the platform.

The imperative need for adopting sign language digital avatars for live-stream and broadcast applications is paramount in our commitment to inclusivity and accessibility. In a world increasingly reliant on digital communication, these avatars serve as indispensable tools in bridging the communication gap for individuals with hearing impairments. By seamlessly integrating sign language avatars into live-stream and broadcast platforms, we not only uphold our responsibility to promote diversity and equal access but also empower millions of deaf and hard-of-hearing individuals to engage fully in the digital landscape. This essential step forward aligns with our dedication to fostering an inclusive society where every voice, regardless of ability, can be heard and understood, enhancing the richness of our shared human experience.

IBC Lip Sync Battle: Bridging the Gap with Sign Language Innovation

The initiative aimed to challenge the misconceptions some have about Deaf sign language users. Not all Deaf sign language users are oral – meaning they don't necessarily use their voices to communicate. The approach provided an opportunity to educate about the diverse ways the Deaf community communicates.

Renowned Deaf actor and sign language expert, Matt Kirby, crafted a sign language version of a weather forecast as a proof-of-concept. This forecast is presented visually entirely in sign language, giving Deaf viewers an equal opportunity to enjoy the weather updates.

The team produced two versions of the signed forecast: one with AI voiceover and another with a voiceover by talented voice actor Darren Altmann. This comparison allows hearing viewers to appreciate the nuances of both presentations.

The project team also developed a digital twin of Matt Kirby, lovingly named Mattheus. AI gives Mattheus a voice. Mattheus could enable red button users to seamlessly switch between a signed forecast and a spoken one. This innovation opens a world of possibilities for Deaf and hearing viewers, allowing them to access information in the format that best suits them.

Impact

The impact of these initiatives could be remarkable if brought to market:

1. Empowering Deaf Presenters: With the success of Matt Kirby and Mattheus, doors open for Deaf presenters to diversify their portfolios. The broadcasting world can recognize the potential of signing talent in various roles beyond weather forecasts.
2. Broadening Exposure to Sign Language: For the first time, hearing viewers are exposed to the beauty and complexity of sign language in this mainstream context. This exposure not only educates but also fosters a sense of inclusivity and perhaps even the desire to learn to sign.
3. Mainstream Integration: The IBC Lip Sync Battle breaks down barriers, integrating sign language into mainstream entertainment. It showcased the Deaf community's talents and brought them to the forefront of pop culture.

4. Conversations About Compensation: The success of Matt Kirby and Mattheus could spark discussions within the broadcasting industry about how signing talent should be remunerated. This could be a significant step toward recognizing the value of sign language expertise.

If the IBC Lip Sync Battle makes waves, it can achieve something beyond entertainment. A powerful reminder that innovation can drive positive change, challenge stereotypes, and foster a more inclusive society. Through the collaboration of technology, talent, and empathy, the IBC accelerator showed that the power of communication knows no bounds, and that it can bridge gaps and bring people together in ways never thought possible.

Conclusions

The scarcity of sign language translators and interpreters make it difficult for deaf people to access written or broadcast content. Machine-generated sign language translation videos have the potential to meet the increasing demand for translations and scalable solutions. Significant progress has been made in the last four years however two key scientific and engineering challenges are yet to be overcome, specifically regarding motion capture and clean-up for sign language performances and fitting translations into the allocated time frame of the source.

AI Sign Language translation has gained momentum in recent years with the emergence of several companies and initiatives, while established translation companies face scalability issues, and large technology companies work on hyper-realistic avatars. There is increasing interest from various industries in applying these solutions.

Generative AI and virtual actors are the two main approaches to this problem. Virtual actors offer several advantages over Generative AI but are not without their challenges. Recent technological advances have made creating high-quality virtual actors both cost and time effective, and it is expected that third-party avatars will match or exceed the realism of Deep Fakes within the next 1-2 years (circa 2024).

Motion capture technology has been used for decades and has become more affordable. There are several approaches and challenges with each, and some key challenges remain to be solved. This remains an area of intense

research and development by the FANGs and by research organisations. We have developed our own hybrid motion capture solution as an interim solution.

British Sign Language (BSL) poses challenges for translation due to the complexity of the language and its differences from English.

Bespoke language models translate source language to the destination language.

The main challenges in AI sign language translation are improving the efficiency of motion capture and fitting translations into available time, with incremental advances focused on improving quality, authenticity, realism, emotions, expressions, body language, and scalability of the platform.

Developing a high-quality speech-to-speech voice cloning model typically demands a substantial amount of training data, including extensive recordings of the target speaker's voice. Acquiring and curating this data can be time-consuming.

Speech-to-speech voice cloning can produce natural and authentic-sounding speech because it involves cloning an actual human voice. Speech-to-speech systems can convey emotions and nuances in speech more effectively because they replicate the tonal and emotional qualities of a human voice. This is important for applications like storytelling, where conveying emotion is crucial.

Voice cloning and lip synchronization are indeed important technologies when it comes to creating digital twins of individuals from the past. These digital replicas can then be 'programmed' to act in new ways, deliver speeches, engage in conversations, or even participate in virtual experiences and simulations. This technology has potential applications in education, entertainment, and various forms of interactive media, where bringing historical or fictional characters to life in a convincing manner is desired.

However, it's important to consider ethical and legal implications when using these technologies, especially when dealing with historical figures, as it raises questions about consent, accuracy, and the potential for misuse. Additionally, the quality and ethical use of such technology should be carefully monitored and regulated to ensure responsible and respectful applications. It was important for this project to get a consent from Maria Calla's estate.

Key learnings and future work

Creating avatars from archive audio and video footage

Extensive research was conducted to gather information about Maria Callas, encompassing a vast array of resources such as photographs, video recordings, and audio clips stored within the broadcaster's archive. This data was strategically used in various stages of our workflow: they were utilized in audio training to enable the algorithm to replicate Maria Callas' distinctive voice; employed to extract images that served as the foundation for constructing a detailed facial mesh; and analysed to gain insights into the performer's movements. Starting from good data is very important when embarking on such endeavours.

At the start of the project, we had concerns about upscaling the archive footage - but we found a really good solution with the photoshop API and the Neural Network-based algorithm. During the project, we thoroughly tested this solution, and we are quite satisfied with the results.

Motion capture and cleanup

As noted above, motion capture and volumetric capture has advanced significantly in the last few years with budget systems providing good results for common use-cases for indie games and films. We note three areas where future work could provide significant benefits.

Firstly, both the Entertainment workstream and the Accessibility workstream developed a hybrid approach combining data from multiple systems to get a workable data set. Maintaining a bespoke hybrid integration is a lot of work. Motion capture suppliers should look to provide a simple integrated off-the-shelf system that accurately captures body, hands/fingers, face, mouth, and tongue rather than expect customers to integrate.

Secondly, motion capture systems still require a lot of manual data clean-up of the animation data before it is usable in production. For this project, we've found that the data cleanup stage takes 3x to 4x longer than the motion capture stage. Most systems don't incorporate a physics model, so data errors that result in a hand penetrating the stomach or finger penetrating the skull aren't automatically corrected although are clearly intuitively wrong to a lay person. The authors call for motion capture systems to include sophisticated human physics models and to embrace AI to assist with

accurately reproducing human motion with the aim of no or very little manual data clean-up of the raw motion data.

Thirdly, we note that at the time of writing the authors are not aware of any hand/finger tracking solution that is suitable for capturing sign language and call on the motion capture industry to cater for this use case. Motion capture gloves tend to have bulky sensors or batteries located on the backs of the hands or wrists, which get in the way when performing signs where one hand touches that point on the other hand.

Some gloves are unable to accurately track what every finger is doing or the precise location of every fingertip, which makes reproducing certain signs impossible. Many gloves get in the way of performing signs where fingers are touching. Commercial marker-less volumetric capture solutions also tend not to accurately capture the exact positions of fingers throughout a performance.

We call on the motion capture industry to develop hand/finger tracking solutions that are suitable for high fidelity capture of the manual elements of sign languages.

Maturity of ecosystem

The volumetric motion capture industry and the AI sign language translation industry are both new and under-developed with a small number of small, highly innovative companies operating in each space. For media and entertainment companies, this represents a significant risk, as business-critical operations could become dependent on the survivability of a specific supplier, with little or no opportunity to second-source or have a contingency plan for a supplier going out of business. Strategies such as providing long-term contracts, building a strong relationship with the supplier, investing in the development of the supplier, collaborating with industry peers and asking Government for assistance with developing the ecosystem, can all help to manage such risks.

Speech-to-speech synthesis

Workflows are available from multiple commercial vendors. The voices are realistic with natural-sounding emphasis, intonation pace, variation and can express a wide range of emotion, accents and delivery styles. Professional voice actors still provide an edge – their use of intonation, speed and pauses improves the interest and understandability of the commentary that is unmatched by synthetic voice. However, synthetic voice technology continues

to experience rapid progress – advances in the last year include the ability to mimic the intonation, emotion and timing of source audio whilst changing the voice – as with other areas of AI, professionals that adopt synthetic voice as a tool in their toolbox will likely have a significant commercial edge over their peers that shun such technology.

Media supply chain automation

In today's rapidly evolving media landscape, a significant portion of the software infrastructure is still not equipped to seamlessly integrate with cloud technologies or harness the power of artificial intelligence. This technological gap has posed substantial hurdles when attempting to establish connections between various systems spanning diverse infrastructures and environments. The challenge, at its core, lies in automating these linkages efficiently.

One of the foremost difficulties encountered in this endeavour is the necessity for asynchronous operations. An illustrative example of this is the process of requesting a mesh, which entails a notable delay before receiving feedback. This delay is primarily attributed to the time it takes to generate the mesh itself, necessitating a seamless orchestration of these asynchronous actions.

Moreover, the evolving landscape has introduced a novel requirement—hybrid automation. This entails the ability to carry out specific actions manually during intermediate stages of the process. Such a need underscores the complexity of integrating legacy systems with modern, cloud-native solutions, demanding a delicate balance between automation and manual intervention to optimize workflow efficiency. In essence, addressing these challenges is pivotal for achieving seamless integration, enhancing efficiency, and future-proofing media-related software solutions.

Avatar modelling and sculpturing

Avatar modelling has indeed seen rapid advances in recent years, particularly with the development of sophisticated 3D modelling tools, as well as advancements in artificial intelligence and computer vision. However, it remains a challenging task, and manual sculpting by skilled artists and 3D modelers remains essential for achieving a high level of detail and realism. The effort required to produce a true facsimile obeys the law of diminishing returns (Pareto Principle).

The issue of model validation remains unresolved. Model validation in avatar creation is critical to ensure that the resulting avatars meet the similarity to the picture of the real person. Validation may involve obtaining feedback from users and testing the avatars in intended applications to ensure that they evoke the desired emotional responses. There is a lack of objective measurement standards that verify visual differences between the avatar and the image of the real person.

Lip sync

Lip sync technologies work well for the general population when used with audio that is clear and has an accent that aligns well to the underlying language models (for instance, General American accents, Indian English accents). Most such technologies struggle with strong regional accents that are under-represented in the source corpus. We've found that Welsh, Scottish (Glaswegian), Yorkshire and Italian accents were particularly problematic.

We have found that even under the above conditions, the best commercially available lip sync for avatars is suboptimal for lip-readers and lip-speakers compared to humans. The reason for this is that there are micro muscular movements of the corners of the mouth, lips, chin, throat, cheeks and around the eyes that are picked up by lip readers and lip speakers from humans but are typically not present in avatar performances.

In the animation industry, the received wisdom is that there are around 15 unique lip patterns used in English speech. For instance, it is believed that the lip patterns for 'bat', 'pat' and 'mat' are identical. But lip speakers and lip readers assert that they can distinguish between these three similar lip patterns with only visual cues. We've found that by combining software, animation and linguistics expertise we can differentiate and reproduce over 500 different lip patterns. Avatars that observe this differentiation are much

more lip-readable, unanimously outperforming the four industry-leading lip-sync solutions in blind user trials.

In the Entertainment workstream, we found that software was less reliable at accurately lip-syncing against singing than against the spoken word. Again, we believe this to be due to the relatively small sample size of sung content in the source corpus used to train the forced alignment models.

We encourage researchers to embrace our findings on micro-muscular movements and the importance of reproducing and differentiating similar but different lip patterns. We encourage further work on forced alignment and lip sync for regional accents and for sung voice.

Lack of broadcast standards

The lack of a standard mechanism for signalling, presenting and switching between signed and unsigned versions of content, is a significant barrier to increasing the quantity of Sign Language on broadcast and streaming services, according to multiple broadcasters. The good news is that there is cross-industry support for creating and implementing a draft specification to this effect. Early work has begun at BT, YouView, Sky and the BBC. DTG, SMPTE, DVB and HbbTV have expressed an interest in contributing. Appendix B provides high-level technical overviews from some project members. Follow-on work should embrace and extend this promising start.

In the Entertainment space, Object-Based Media (OBM) will play a crucial role in enabling interactive storytelling. Overall, the standardisation of OBM looks to be progressing well however the intersection of OBM and the metaverse (being the catch-all term for extended reality (XR) and connected virtual worlds) is under-explored at the time of writing and should be a focus of future work.

Focus groups

Organisations advocating for the d/Deaf community are often stretched thin, addressing multiple sectors to improve access and services for sign language users. This could be a factor in why our previous attempts to engage with these organizations had limited outcomes. We extend an open invitation to d/Deaf organizations to collaborate with the Media and Broadcast industry. Together, we can create a safe space for meaningful discussions that can truly shape results. For future initiatives, we propose establishing a focus group of Deaf individuals from the beginning, and we recommend allocating a budget to fairly compensate these contributors at market rates for their valuable time.

Commercial frameworks

It's crucial that Deaf organizations and Deaf talent actively participate in a conversation to build and shape standards for IP management, licensing, and distribution. Their insights will help shape fair and effective standards, ensuring that sign language content is both accessible and respectful of cultural nuances.

The evolving legal and commercial landscape

At the time of writing, both the Writers Guild of America (WGA) and the Screen Actors Guild and American Federation of Television and Radio Artists (SAG-AFTRA) 2023 members were striking. The legal and contractual landscape is likely to evolve significantly over the next few years, as the strikes are resolved, the action points implemented, and follow-on work is identified and negotiated.

The authors believe that the outputs from the resolution of the Writers and Actors strikes can inform the commercial frameworks discussion in the previous section.

Identification and remuneration

Talent identities such as the DOI HAND ID are likely to form part of that landscape, particularly in areas such as provenance, identification, verification, accounting, remuneration, security and automation.

Summary

In summary, a project like this requires a careful balancing act between legal obligations, ethical considerations, and practical needs. Collaborating with legal professionals, ethicists, and representatives from the deaf community will help ensure that the project is carried out responsibly and respectfully.

This document was created by the human champions and participants of the Synthetic Humans for Entertainment and Accessibility project. A ‘version 0’ outline draft of this document was created with AI assistance.

Appendix A – Policy Statements

The project team produced an AI policy statement and an Environmental policy statement suitable for micro and small production facilities that haven't yet adopted their own policies in these areas. The team has placed these policies in the public domain, so SMEs are free to adopt, extend and use as they see fit.

Please retain the Public Domain logo when publishing policies verbatim and please consider releasing any derivative documents to the public domain for the benefit of customers, suppliers, everyone working through the product value chain, and society.

The following policy is intended as a starting point for micro and small production facilities that don't have an Environmental Statement and are looking to create one but don't know where to start. It is important that organisations adapt this document to their own needs and treat it as a living document that will evolve over time.

Organisations are particularly encouraged to review the B-Corp[42], BAFTA (Albert)[43], Ecologi[44] and ESGmark[45] sustainability criteria.

Environmental Statement

<Company> recognises that maintaining an awareness of environmental issues is synonymous with good business practice and corporate social responsibility in minimising any potentially negative impacts on our environment. Whilst we do not consider the nature of, or outputs from, our business to contribute to significant environmental detriment, our aims are summarised by the three R's:

Review

– all aspects of our business on a continuous basis

Reduce

– unnecessary waste from every part of our business

Recycle

– as much as possible from our daily operations

We are committed to compliance with all relevant environmental legislation, regulations, and codes of practice and to work closely with clients and other stakeholders to achieve best practice within the industry sectors in which we operate.

Environmental policy, like all other *<Company>* policies, is not considered as stand-alone but is woven into the fabric of our day-to-day business processes to ensure familiarity and focus towards continual improvement in this important area of our work. Our policy is characterised by the following key corporate guidelines:

Make efficient use of all resources used by the company as part of its daily operations by conserving energy and water, minimising waste and recycling equipment and materials where practically possible.

Select and use environmentally friendly and recycled materials wherever they can be commercially justified.

Foster a transport policy which encourages optimised vehicle efficiency through regular maintenance and driver guidance, whilst also encouraging the reduction of unnecessary journeys where other modes of communication may be easily available.

---

[42] https://www.bcorporation.net/en-us/

[43] https://wearealbert.org/

[44] https://ecologi.com/

[45] https://www.esgmark.co.uk/

Encourage members of staff to provide input to the development of company policy and feedback to the management regarding the company's environmental performance.

<Company> makes heavy use of cloud services. This policy requires *<Company>* to secure such services from companies with a progressive zero carbon policy. Google, Amazon, Microsoft, IBM, and Facebook all have policies that meet this criterion.

For instance, in Google has been carbon-neutral since 2007 and in September 2020 Google announced that it had compensated for all the carbon it had produced since it was created. In January 2021, Microsoft said it will reach 100% renewable energy by 2025, becoming carbon negative by 2030 and will eliminate all past carbon emissions by 2050. Amazon, Facebook and IBM all have similar ambitions.

This policy will be subject to review on an annual basis, taking into account any changes in legislation, identified client specification and operational experience. Any updates will be disseminated to all staff, and any material changes will be highlighted and subject to further discussion with all stakeholders.

<Company> will make this policy available when requested to any and all interested parties and will continue to work with all stakeholders to improve the way we manage our approach to environmental responsibilities as time proceeds.

Sustainable Production

As part of our Environmental Policy, we've adopted the following policies and practices around sustainable production across offices, production office and studio facilities:

- Use of LED lighting or other very low energy lighting where possible.
- Zero-waste to landfill policy.
- Use of mains as the primary power source. No use of diesel or petrol generators
- Mains power sourced from a 100% renewable sourced energy tariff.
- Use of lighting sensors, visual reminders on energy savings and waste management
- Production office produces production documents electronically by default. Only printed out on paper when explicitly requested by individual team members, for only them.
- Use of on-line magazines, newspapers, emails by default.
- Copier paper is sourced from verified sustainable sources (EFC, PEFC, EU Ecolabel or other paper made from recycled material).
- Where batteries are used, they are rechargeable batteries.
- *<Company>* productions do not use physical props, sets, costumes, make-up, SFX, etc.
- Use of re-usable water bottles for cast and crew.
- Where catering is provided, all meals are vegetarian/vegan by default, individuals have to opt-in for alternative dietary requirements. Catering composts and recycles all food waste including packaging. Single-use products are not used. Food is sourced from suppliers who prioritise based on their environmental impact.

We welcome any feedback that will help us improve this policy for the benefit of our customers, suppliers, everyone working through our product value chain, and society.

The following policy is intended as a starting point for micro and small production facilities that don't have an AI Statement and are looking to create one but don't know where to start. It is important that organisations adapt this document to their own needs and treat it as a living document that will evolve over time.

Organisations are particularly encouraged to review the links provided below.

AI Statement

<*Company*> recognises that whilst AI has the potential to significantly improve our lives (e.g. improving healthcare, production efficiencies, mitigating and adapting to climate change) maintaining an awareness of the potential risks that AI poses (such as opaque decision-making, discrimination, intrusion into family and private life and being exploited by criminals) is synonymous with good business practice and corporate social responsibility in minimising any potentially risk and negative impacts on our lives.

To safeguard that our technology is trustworthy and developed in accordance with globally recognised ethics standards that address potential adverse impacts on Human Rights, we subscribe to The Turing Institute's guidelines for responsible innovation, as referenced by the UK Government[46]:

- **ethically permissible** – consider the impacts it may have on the wellbeing of affected stakeholders and communities.
- **fair and non-discriminatory** – consider its potential to have discriminatory effects on individuals and social groups, mitigate biases which may influence your model's outcome, and be aware of fairness issues throughout the design and implementation lifecycle.
- **worthy of public trust** – guarantee as much as possible the safety, accuracy, reliability, security, and robustness of its product.
- **justifiable** – prioritise the transparency of how you design and implement your model, and the justification and interpretability of its decisions and behaviours.

We also subscribe to the EU's ethics guidelines on Trustworthy AI[47].

Briefly, the guidelines state that Trustworthy AI shall be:

- **lawful** – respecting all applicable laws and regulations.
- **ethical** – respecting ethical principles and values.
- **robust** – from a technical perspective, whilst taking into account its social environment.

<*Company*> complies with all relevant AI legislation, regulations, and codes of practice and is committed to working closely with clients and other stakeholders to achieve best practice within the industry sectors in which we operate.

Whilst we do not consider the nature of, or outputs from, our business to contribute to significant social-economic detriment, our aims are to actively review and manage potential risks.

[46] https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety

[47] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

We've put mechanisms in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. We also provide an adequate and accessible route of redress and systems to ensure continuous monitoring and improvement of the same.

AI policy, like all other <*Company*> policies, is not considered as stand-alone but is woven into the fabric of our day-to-day business processes to ensure familiarity and focus towards continual improvement in this important area of our work. Our AI policy is characterised by the following seven key corporate AI guidelines:

- **Human agency and oversight**
- **Technical robustness and safety**
- **Privacy and data governance**
- **Transparency**
- **Diversity, non-discrimination and fairness**
- **Societal and environmental wellbeing**, and
- **Accountability**

These seven guidelines are the same as those that form the EU's AI whitepaper, published February 2020[48].

This policy will be subject to review on an annual basis, taking into account any changes in legislation, identified client specification and operational experience. Any updates will be disseminated to all staff, and any material changes will be highlighted and subject to further discussion with all stakeholders.

<*Company*> will make this policy available when requested to any and all interested parties and will continue to work with all stakeholders to improve the way we manage our approach to AI responsibilities as time proceeds.

Further reading:

- UK Government AI Industrial Strategy (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf)
- UK Government Office for Artificial Intelligence (https://www.gov.uk/government/organisations/office-for-artificial-intelligence)
- Global Partnership on Artificial Intelligence (https://www.gov.uk/government/publications/joint-statement-from-foundingmembers-of-the-global-partnership-on-artificial-intelligence/joint-statement-from-founding-members-of-the-globalpartnership-on-artificial-intelligence)

[48] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

- The Alan Turing Institute's guidance on AI ethics and safety (https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf)
- UK Government guidance on understanding AI Ethics and Safety https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety
- EU Ethics guidelines for trustworthy AI https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
- European Commission White Paper on AI: A European approach to Excellence and Trust https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

Future versions of this policy are likely to incorporate and reference:

- the IEEE standards on Ethically Aligned Design (https://standards.ieee.org/wpcontent/uploads/import/documents/other/ead_v2.pdf)
- the OECD principles on Artificial Intelligence (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449)
- the Partnership on AI's (https://partnershiponai.org/paper/shared-prosperity/) Guidelines for AI and Shared Prosperity (https://partnershiponai.org/paper/shared-prosperity/)
- the EU's proposed AI Act (https://artificialintelligenceact.eu/)
- the UK Government's white paper and proposed regulation on AI (https://www.gov.uk/government/publications/ai-regulation-apro-innovation-approach)

We welcome any feedback that will help us improve this policy for the benefit of our customers, suppliers, everyone working through our product value chain, and society.

AI Statement  : Public domain

Appendix B – Platform Descriptions

Some members of the project team produced a 1- or 2-pager high level technical overview of how they currently (or would) implement signalling, presentation of choice and end-user switching between BSL and non-BSL versions of a programme. today and how they'd ideally like it to work within their own system in, say, 3 years' time, with the intention of sharing the documents with industry partners, to stimulate further discussion.

The aim was then to then document:

- Areas of convergence between all media & broadcast companies – these are areas that are candidates for eventual standardisation
- Areas of divergence – these could be documented in a standard, in a non-normative section, perhaps with guidelines or notes on best practices

Where permitted, the documents are included in Appendix B to provide a foundation for future work in this area.

High Level Technology Overview (BT):
BSL Signalling, End-user presentation and Switching

Version 0.2 MR (BT)

Revision History

Version 0.1	*02/06/2023*	*Mark Riley*	*Initial version for peer input and discussion*
Version 0.2	*04/07/2023*	*Mark Riley*	*Enhanced based on further thought and peer input*

Introduction

This design note outlines the High Level Technology direction that BT intends to take for British Sign Language (BSL) production, end-user presentation, signalling and switching for sign-supported VoD content. By setting out our aims and roadmap, we intend to influence other UK broadcast players and device manufacturers to achieve a greater level of BSL presentation flexibility for end-users.

Background Brief

As technical architect for Consumer TV Access Services in BT, I was asked to provide a view of our intended approach to signalling, presentation of choice and end-user switching between BSL and non-BSL versions of a programme.

IBC Synthetic Humans Project are asking each broadcaster and streaming service to provide a high- level technical overview (1- or 2-page high-level technical overview that can be shared with industry partners, to stimulate further discussion) of how they currently (or would) implement this today and how they'd ideally like it to work within their own system in, say, 3 years' time.

Their aim is to then document:

- Areas of convergence between all media & broadcast companies – these are areas that are candidates for eventual standardisation
- Areas of divergence – these could be documented in a standard, in a non-normative section, perhaps with guidelines or notes on best practices?

Overall Approach

Our aim is to use Object-Based techniques to make access to BSL-signing within VoD programming more dynamic (can switch signer on/off during programme playback) and give the end-user control over presentation (positioning, size and even the device upon which the signer is presented). Based on feedback from a variety of BSL users, we believe that adding such flexibility will greatly improve the end-user experience and drive further use of the platform. The ability to simultaneously present subtitles, adjusting their use of screen real-estate to compliment rather than clash with the signer (if present) may also be of value to a worthwhile sub-set of users.

The traditional approach of producing a separate pre-rendered signed version of each VoD asset, while capable of delivering results, seems inflexible and archaic – particularly given the trend towards spatial computing, where treating individual elements of a visual composition as separate but linked objects will become standard practice.

Contribution Approach

We would much prefer to receive object-based BSL mezzanine from content partners rather than a pre-composited signed version of an asset. This would gives much greater flexibility over time to be able to generate signed content suited to the capabilities of the display device, be that in a traditional 2D TV viewing environment or in a multi-device or spatial 3D context.

The need for a standardised approach to Object-Based signing production, and agreeing an exchange format for mezzanine-level deliverables between UK broadcast industry peers (and beyond) would greatly increase the likelihood of widespread adoption for a more flexible approach, and encourage device manufacturers to support agreed object-based distribution formats.

We propose that a mechanism for mutual exchange of BSL-signed content be established, potentially via an existing forum (e.g., DTG/TODIF Access Services) that sets out a Memorandum-of-Understanding (MoU) across Content and Service providers to allow a programme (be that UK or rest-of-the-world originated) to be BSL-signed once, but that signing production to then be re-used by any UK VoD or linear services carrying that content with minimal additional cost – providing the greatest benefit to end-users of BSL signing services, and allowing available access investment to be spent on widening the breadth of programming with BSL rather than duplicating effort signing content which already has BSL interpretation/presentation.

Distribution Approach

The use of HEVC (or a similar next-generation video CODEC) which supports encoding with an Alpha transparency layer[49] would seem key to defining the distribution format. In our tests, an HD-resolution capture of a signer, either as 16:9 landscape or, preferably 16:9 portrait filling the majority of the frame. This should have a digital transparent background - derived from green-screen capture of the physical signer in our case, or natively transparent when using a computer-rendered signer. This has proven to be a versatile format, allowing us to superimpose the signer onto both HD and UHD content. For UHD-2 (8K) content, signer capture at UHD is required to give respectable results.

Presentation Approach

The ability of the VoD playback device to dynamically superimpose the signer (which is a full-motion HD video layer) on top of moving video in a synchronised manner is key to providing an object-based service. This is typically achieved in STB devices using a graphics plane which sits above the video plane in the presentation stack – often such a graphics plane is used for subtitle overlay, but in this case the plane needs to be real-time video capable rather than just for text/graphics.

In modern systems, which employ an HTML-5 browser (or similar) for graphics plane rendering, video playback of the BSL signer via this browser technology may be possible, but in many cases may not be hardware accelerated. Working with device partners to achieve the lowest-impact synchronised signer playback (in parallel with main video decode) is a key part of our strategy;

---

[49] Transparency | AVPro Video - Documentation (renderheads.com)
Not many video codecs have native support for transparency / alpha channels, but is supported by HEVC.

this also opens the door to achieving the kind of flexible presentation we would like to achieve, but requires dynamic use of hardware video scaling engines, transparent overlay and background underlay, rather than just a simple windowed P-in-P (Picture-in-Picture) approach.