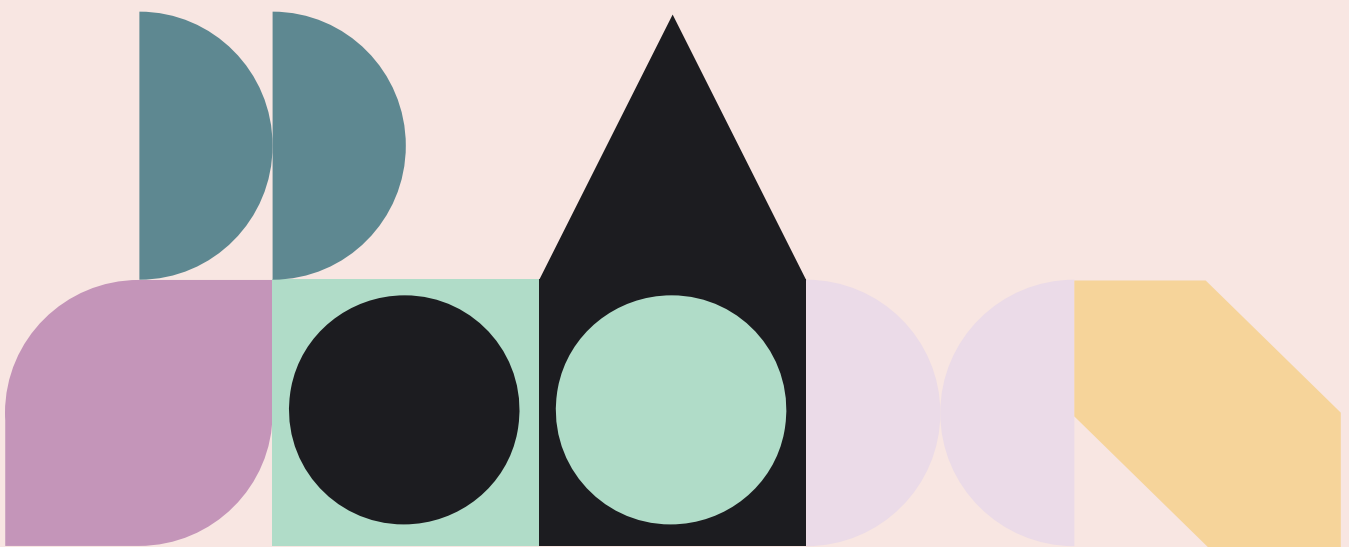


Llais V2

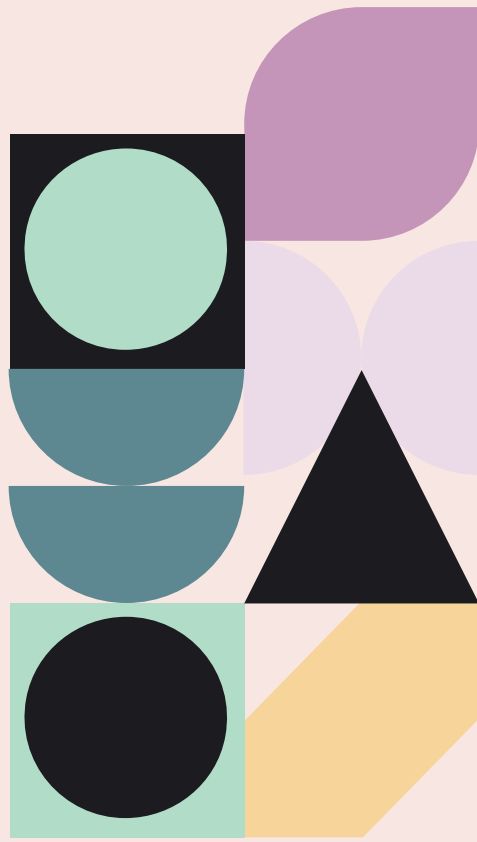
Advancing AI-Driven Translation, Lip-Sync & Dubbing

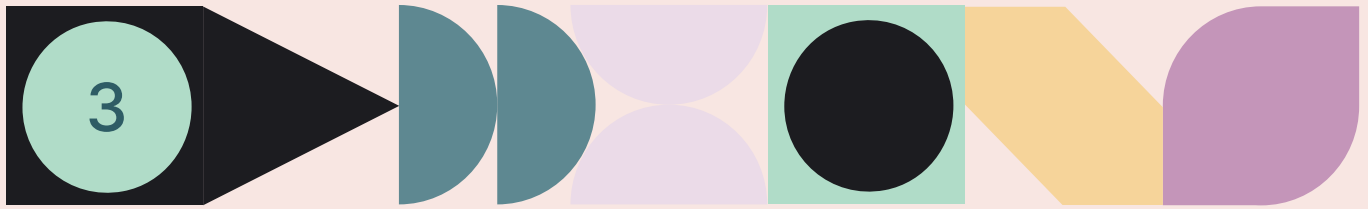
A Proof of Concept Report by
Aled Jones



2026

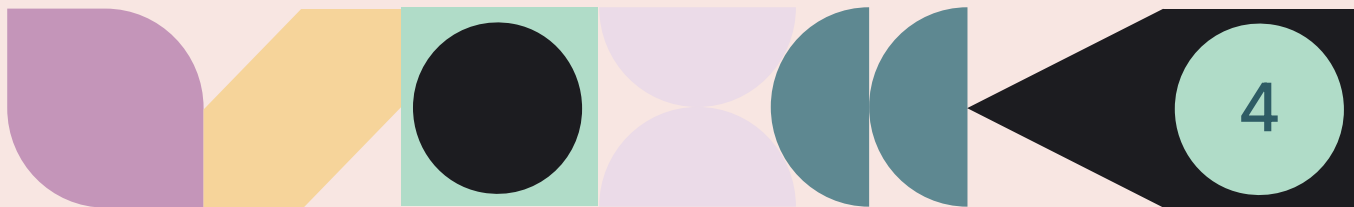
media
cymru × pdr.





contents

4	Executive Summary
6	Introduction
7	Project Objectives
8	Project Team
10	Deliverables & Timeline
12	Development of the Production Pipeline
16	Engagement with Industry Stakeholders
18	Intergration with Existing Workflows
19	Testing with Broadcast Content
24	Key Challenges and Learnings
25	Commercial Potential and Future Development
26	Conclusion
28	Showcasing Llais V2
30	Glossary
32	Appendix



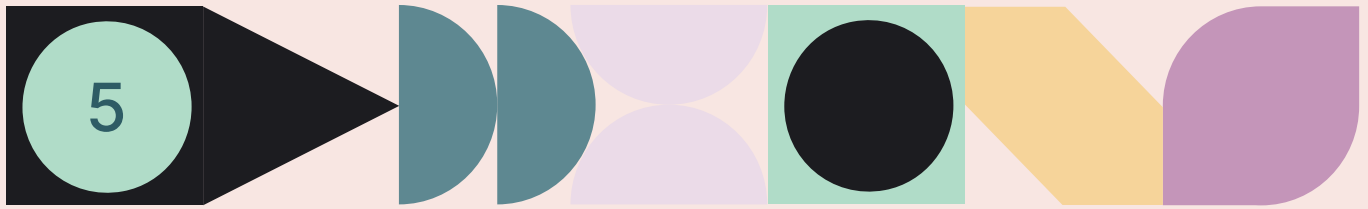
executive summary

Llais V2 builds on the initial Llais project, moving from early-stage research into the practical development and testing of an AI-driven lip-syncing solution for media localisation. The focus of this phase was not only to refine the underlying technology, but to understand how it could operate within real production environments and develop into a commercially viable service.

At its core, the project set out to design a production pipeline that could integrate AI processes - such as transcription, translation, audio alignment and lip synchronisation - into existing post-production workflows. A key consideration throughout was how to automate elements of the process without displacing the creative and technical professionals who underpin the industry. Instead, the approach taken positions AI as a tool that enhances and supports translators, performers and editors, rather than replacing them.

Testing the system on broadcast material proved essential. It allowed the team to move beyond controlled conditions and assess how the technology performs in practice. The results were encouraging, particularly where high-quality footage and carefully prepared training data were available. At the same time, the testing made clear that outcomes are highly dependent on the characteristics of the source material. Factors such as lighting, resolution, shot composition and movement all influence the quality of the lip-sync, and in many cases, human intervention remains necessary to achieve a broadcast-ready result.

Engagement with industry stakeholders was central to the project. Conversations with post-production specialists, production companies and representative bodies highlighted both the appetite for innovation and the need for caution. There is clear interest in tools that can streamline workflows and open up new possibilities for content distribution, particularly for Welsh-language media reaching international audiences.

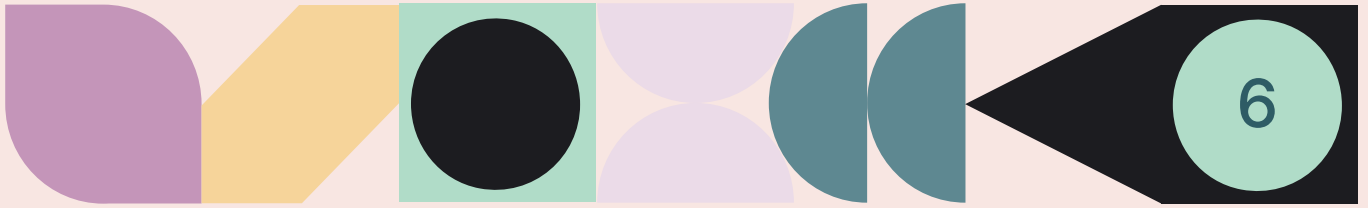


However, these opportunities sit alongside important ethical and legal considerations, including performer consent, ownership of likeness, and fair remuneration. These discussions have directly shaped the project's direction, ensuring that the technology is developed in a way that aligns with industry values and expectations.

Another consistent theme was the importance of compatibility with established workflows. For the technology to be adopted, it must fit naturally within the tools and processes already used across the sector. This includes working with standard editing platforms and formats, and allowing content to move easily between systems without adding complexity. Encouragingly, stakeholders responded positively to the idea of a simple, interface-led approach, particularly where it could support tasks such as dialogue replacement or scene correction without requiring costly reshoots.

From a commercial perspective, Llais V2 has demonstrated that there is a credible pathway towards a scalable service. The potential applications extend beyond straightforward localisation, offering value in post-production efficiency and creative flexibility. While further development is needed - particularly in improving automation and handling more complex footage - the project has established a strong foundation on which to build.

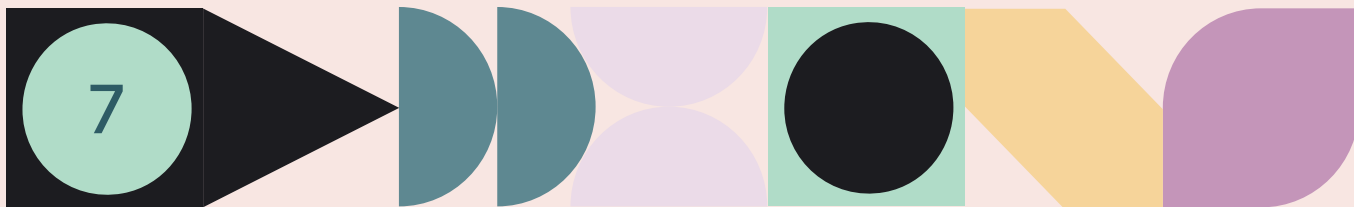
Llais V2 shows that AI-driven lip-syncing can play a meaningful role in the future of media production, provided it is implemented thoughtfully. The project has not only advanced the technical capabilities of the pipeline, but also clarified the practical, ethical and commercial context in which it must operate. In doing so, it sets out a clear direction for the next stage of development and the wider adoption of the technology within the industry.



introduction

Llais V2 builds directly on the earlier Llais project, which explored the potential of AI-driven lip-syncing as a way of making Welsh-language content more accessible to wider audiences. While the initial phase focused on feasibility and ethical considerations, this second phase shifted towards practical application. The emphasis moved from asking whether the technology could work, to understanding how it might function within real production environments.

This meant developing not just the technology itself, but the surrounding processes that would allow it to be used in a professional context. From the outset, the intention was to create something that could integrate into existing industry workflows rather than sit outside them. At the same time, the project remained grounded in the cultural and linguistic context that inspired it, with a particular focus on supporting the reach of Welsh-language media.



project objectives

The project was guided by a clear set of ambitions. It sought to develop a production-ready pipeline for AI-assisted lip-syncing, to test that pipeline using real broadcast material, and to explore whether the approach could form the basis of a viable commercial service.

Alongside the technical goals, there was an ongoing commitment to understanding how such a tool would be received by the industry. This included engaging with stakeholders, identifying potential barriers to adoption, and ensuring that the system was designed in a way that complemented, rather than disrupted, established creative roles.

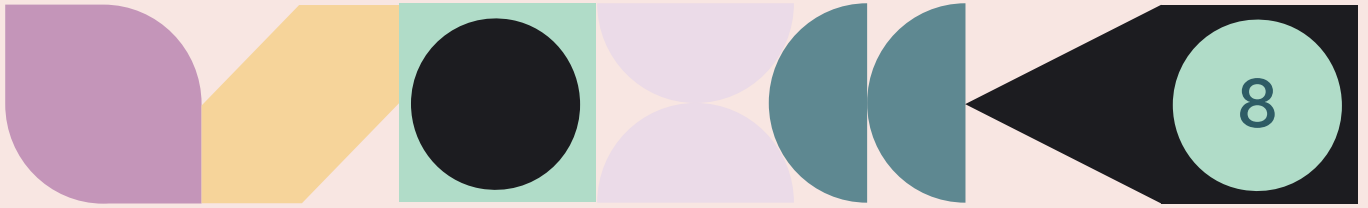
In practical terms, the work focused primarily on Welsh-to-English localisation, while also considering how the same approach might be extended to other languages and markets in future.

Objectives

- *Develop a production-ready AI lip-sync pipeline*
- *Automate key stages of the localisation process*
- *Test the system using real broadcast material*
- *Engage stakeholders to assess industry readiness*
- *Explore commercial models and pricing structures*

Scope

- *Welsh-to-English localisation (primary focus)*
- *Integration with existing post-production workflows*
- *Use of both archive and newly broadcast material*



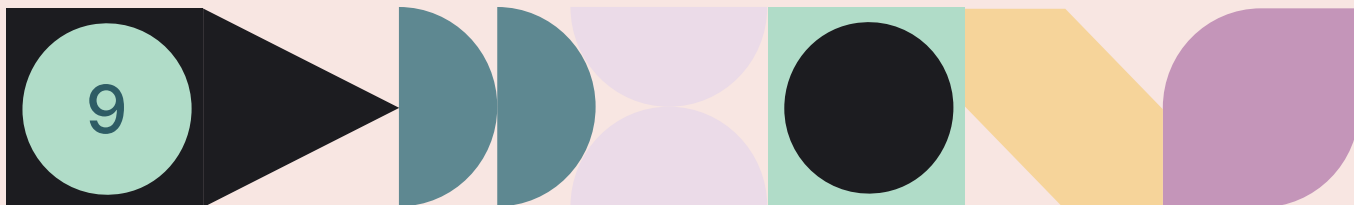
project team



Project Lead
Aled Jones

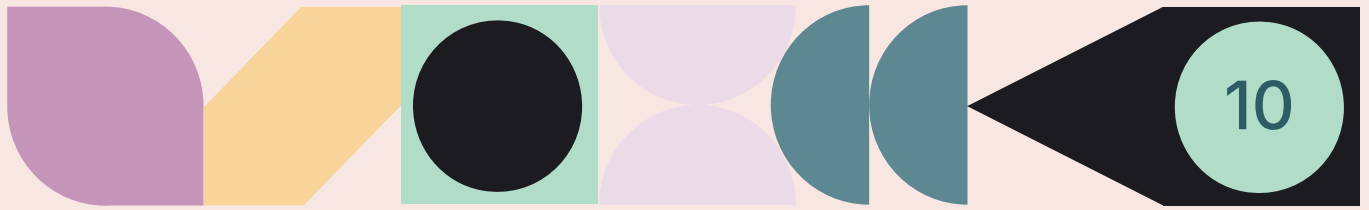
Aled Jones serves as the director of Y Pod Cyf, specialising in Welsh language digital initiatives. Throughout his career, he has managed and developed innovative projects that have contributed significantly to shaping the Welsh digital landscape.

Aled led Llais, a Media Cymru seed funded project, and sought to investigate the potential impacts and ethical considerations of using AI lip-syncing tools for producing Welsh-to-English video content, aiming to make Welsh-language content accessible to broader audiences. With over twenty years of experience in the Welsh media industry, Aled was able to leverage his extensive network to engage industry stakeholders and gain insights into AI's role in the creative sector.



Technical Partner
Robin Moore

Robin Moore, Director at Shwsh and Innovation Consultant with Media Cymru, serves as a technical partner in the Llais V2 project, offering AI expertise.



deliverables & timeline

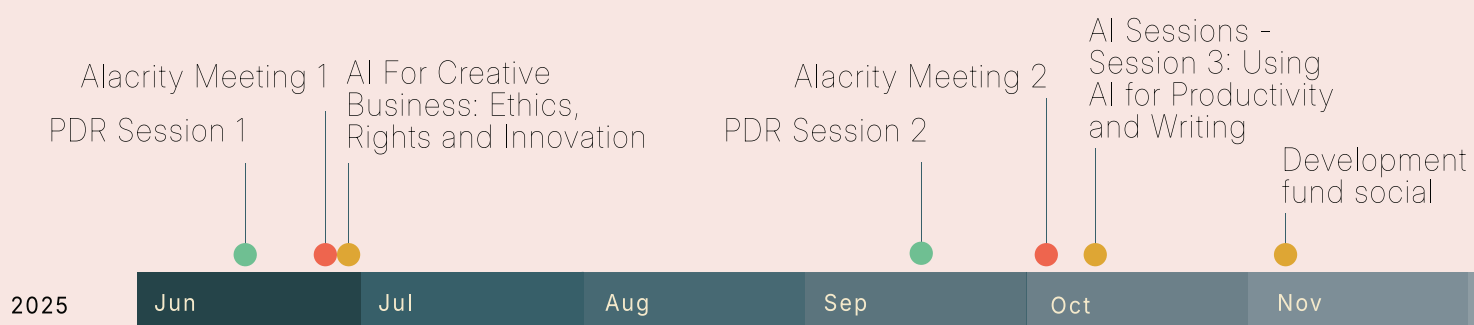
The project was structured around a series of defined deliverables, reflecting its progression from planning and development through to testing and evaluation.

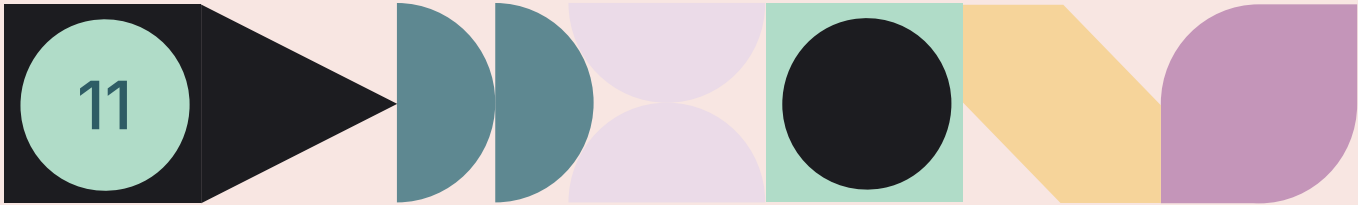
Early stages focused on refining the pipeline specification and updating project plans, followed by pilot testing and the development of a clearer business model.

The final outputs bring together the technical findings, commercial considerations and overall assessment of the project's potential.

Summary of Deliverables

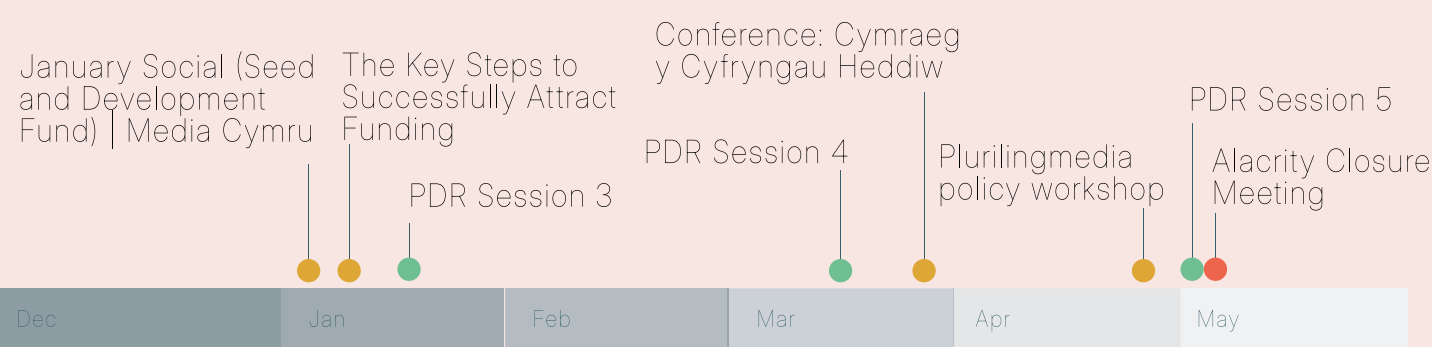
Ref	Title	Due Date
D1	Pipeline Specification	Sep 2025
D2	Project Plan Update & Contracts	Oct 2025
D3	Pilot Tests & Results	Dec 2025
D4	Project Plan Update	Jan 2026
D5	Business Model Development	Mar 2026
D6	Proof of Concept Report	May 2026
D7	Final Claim & project Update	May 2026

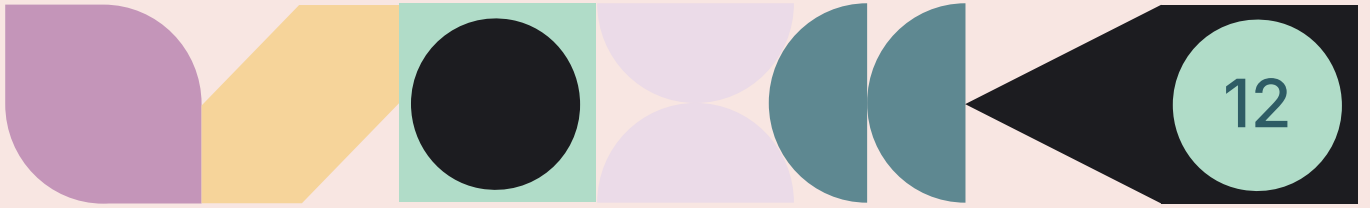




Summary of Milestones

Ref	Title	Due Date
M1	Project Start	Jun 2025
M2	Investigate & compare business models (alongside pilots)	Sep 2025
	Development of pipeline design/feasability	
M3	Milestone meeting 1 with Media Cymru	Oct 2025
M4	Investigate & agree programme partnerships for pilots	Nov 2026
	Development of Ingest tools	
	Improve Welsh support of core models	
M5	Pilots & Testing	Dec 2025
M6	Milestone meeting 2 with Media Cymru	Jan 2026
M7	Investigate & compare business models (alongside pilots)	Apr 2026
	Balance AI automation versus Human oversight & performance	
	Improve outputs and review tools	
M8	Develop Final Report	May 2026
M9	Closure meeting with Media Cymru	May 2026
M10	Project end	May 2026





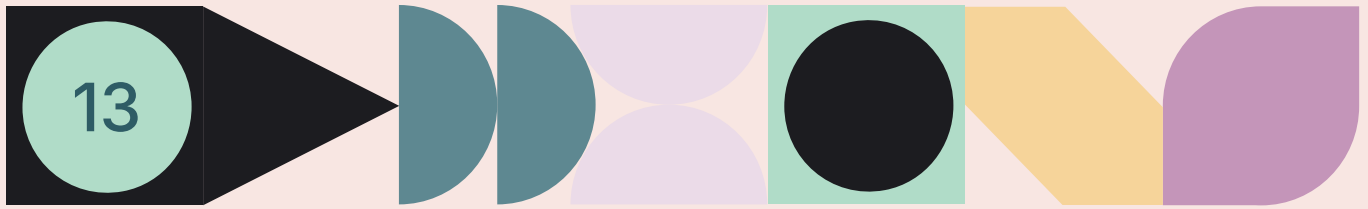
development of the production pipeline

A significant portion of the work centred on developing a pipeline that could support the full localisation process from initial ingest through to final output. Rather than treating lip-syncing as a standalone step, the project approached it as part of a broader workflow.

This began with the ingestion of media, including video, scripts and audio elements, followed by automated analysis. Tools such as transcription and speaker identification were used to break content into manageable, timecoded segments. These segments then formed the basis for translation and script adaptation, allowing dialogue to be reshaped in a way that would better fit the timing and rhythm of the original performance.

From there, the process moved into audio creation, either through traditional ADR workflows or AI-generated voices, before progressing to alignment and lip-sync processing. The final stages focused on quality control and ensuring that outputs could be reintegrated into existing editing environments without disruption.

Throughout this process, it became clear that preparation is critical. The quality of inputs - whether in terms of footage, audio or metadata - has a direct impact on the quality of the final result.



Key Development Areas

Client-Facing Portal

A platform enabling upload of video, scripts, audio stems, and training material.

Automated Transcription & Segmentation

AI tools used to align scripts with footage using timecodes.

AI Lip-Sync Training

Refinement of models to match actor lip movement accurately.

Editing & Quality Control

Post-processing workflows to ensure natural and broadcast-ready outputs.

Ethical Framework

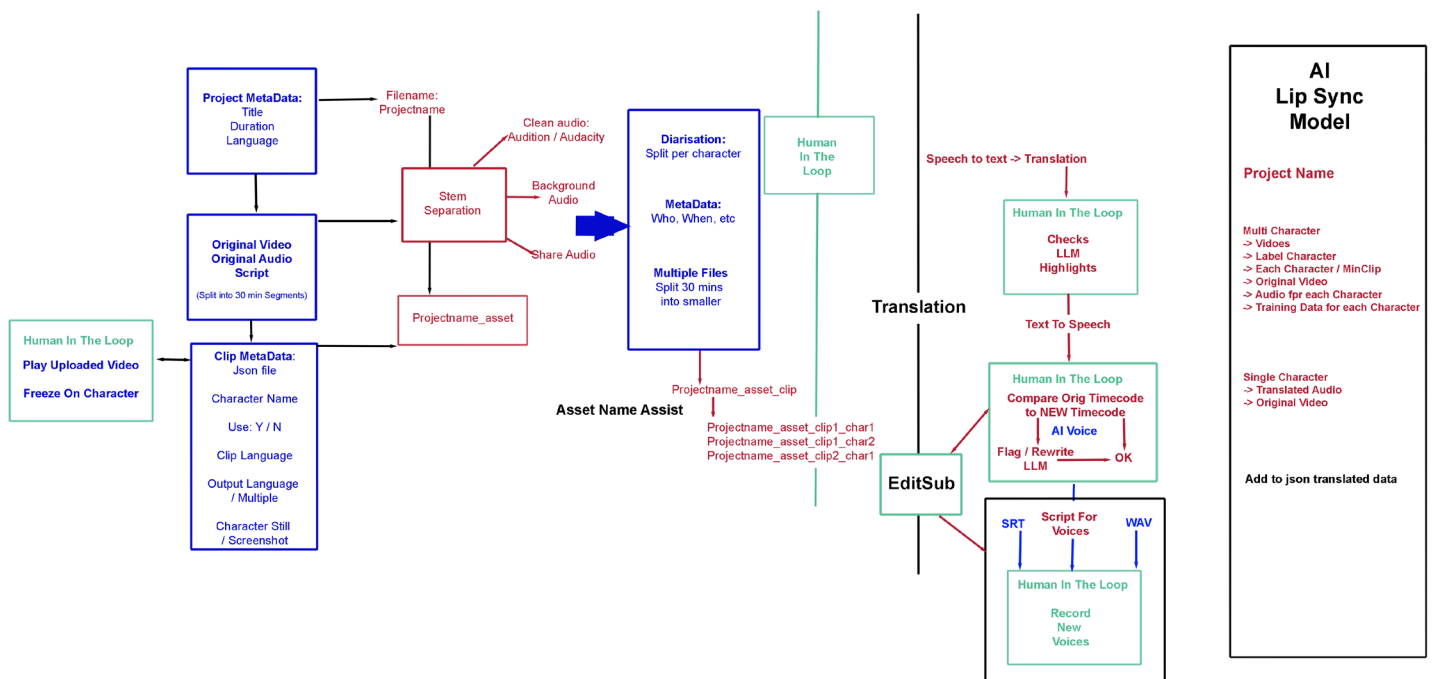
Ensing safeguarding of creative roles and alignment with industry standards.

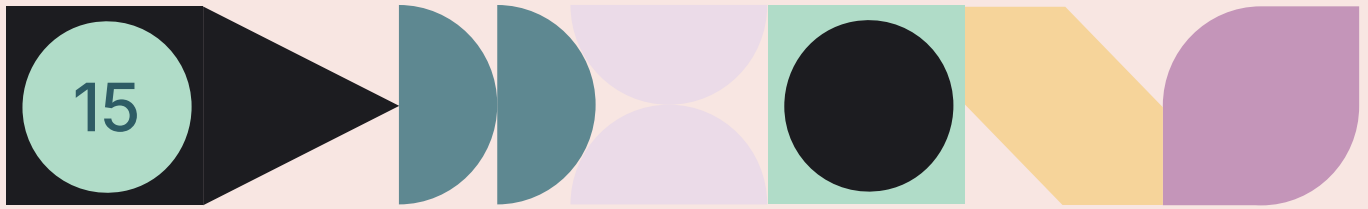
Technical Workflow & Pipeline Development

The Llais V2 pipeline consists of several modular stages:

Core Pipeline

1. Ingest & Upload
2. Media Analysis (transcription, diarisation, segmentation)
3. Manual Review Interface
4. Translation & Script Adaptation
5. Audio Creation (ADR or AI voice)
6. Audio Alignment
7. Lip-Sync Processing
8. Final QA & Delivery





Technical Summary

The technical R&D phase focused on addressing key bottlenecks in the AI dubbing and lipsync production workflow.

Three main challenges were identified:

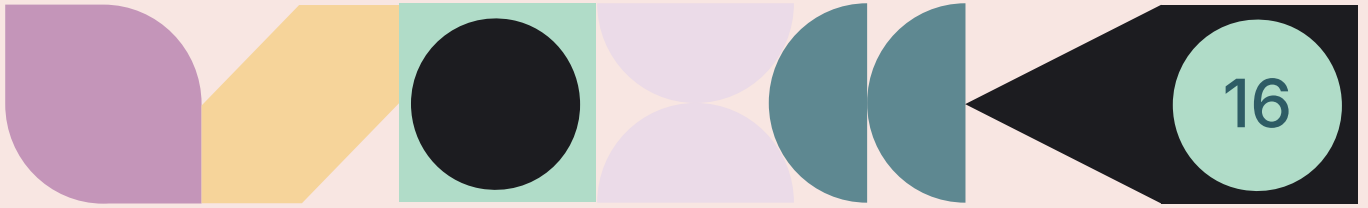
- the time required for transcription, translation, and dialogue alignment;
- the accuracy of Welsh-language transcription;
- the efficient delivery of production assets required for AI model training.

To improve localisation workflows, a tool called DubTool was developed. DubTool integrates transcription, subtitle editing, speaker management, translation, and AI-assisted duration testing within a single interface. By allowing translated dialogue to be evaluated against timing constraints earlier in the process, the tool reduces the need for costly revisions during later alignment stages and streamlines preparation of dubbing scripts.

Research into Welsh-language transcription compared multiple AI transcription systems, including Speechmatics, AssemblyAI, and a Welsh-trained Whisper model. Results showed that while the Whisper model generally achieved higher accuracy, each system exhibited distinct strengths and weaknesses. Further research demonstrated that combining outputs from multiple transcription models using a Large Language Model (LLM) could significantly improve transcription quality. This approach is planned for future integration into DubTool.

The project also investigated challenges associated with obtaining sufficient training material and production assets for AI lipsync generation. To address these issues and simplify client delivery workflows, a prototype Llais Ingest Portal was developed. The portal enables secure upload, validation, tracking, and preview of editorial delivery packages while generating structured metadata for downstream processing. The system supports existing production workflows and provides a foundation for more efficient asset management and lipsync production in future deployments.

(* Full details in the Technical R&D Summary)



engagement with industry stakeholders

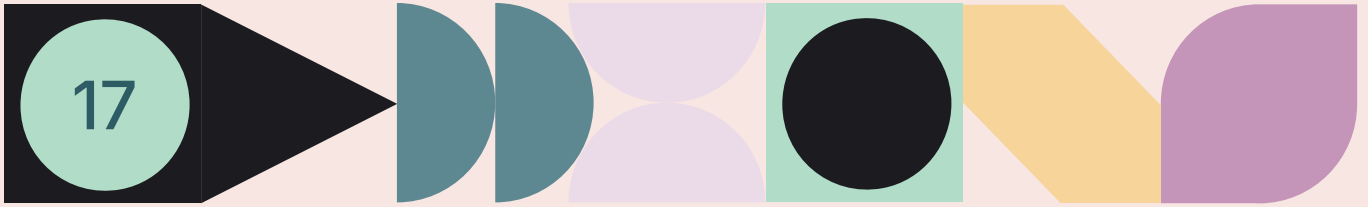
Engagement with industry stakeholders was a central part of the Llais V2 project and provided valuable insight into both the opportunities and challenges associated with adopting this technology.

Engagement was undertaken through a series of informal discussions and interviews with a broad cross-section of the sector, including the Welsh language broadcaster S4C, independent production companies working in Welsh-language broadcasting, drama television directors, writers and producers, post-production professionals, audio dubbing experts, facility managers, and union representatives from both UK-wide and Welsh perspectives.

These discussions highlighted a strong interest in tools that could streamline workflows and reduce the need for reshoots. In particular, the potential to correct dialogue or adapt content late in the production process was seen as a significant advantage by production and post-production professionals.

At the same time, conversations with representative bodies brought important ethical considerations into focus. Questions around consent, ownership of likeness, and fair remuneration are not peripheral issues but fundamental to how this technology can and should be used. The project has taken these concerns seriously, recognising that long-term adoption depends on building trust as much as technical capability.

Overall, engagement with stakeholders reinforced both the opportunities and the concerns surrounding AI adoption, underlining the importance of continued dialogue alongside.



Ethical & Legal Considerations

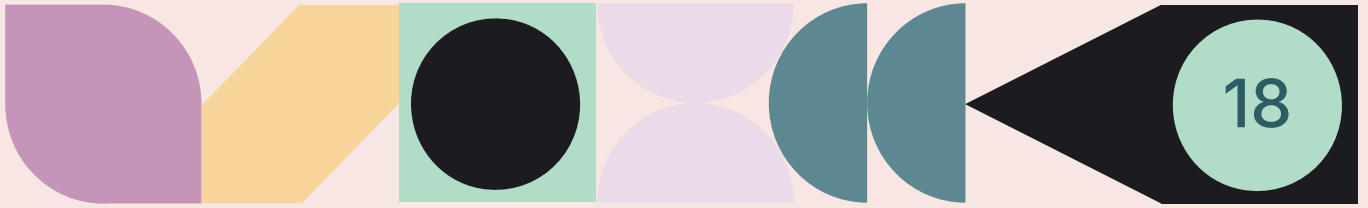
Engagement with representative bodies emphasised:

- Performer consent and data rights
- Ownership of likeness (face, voice, identity)
- Moral rights and attribution
- Fair remuneration and transparency

AI is not yet explicitly regulated in UK law, but existing copyright and data protection frameworks apply.

Industry Feedback Highlights

- Strong interest in post-production integration
- Preference for simple upload-based interfaces
- Demand for broadcast-quality outputs
- Concerns around:
 - Performance authenticity
 - Ethical use of likeness
 - Workflow compatibility



integration with existing workflows

One of the clearest messages from stakeholders was that any new tool must fit within existing production environments. The industry relies on a well-established set of software tools for editing, grading, visual effects and audio post-production, and there is little appetite for solutions that require entirely new ways of working.

As a result, the project explored how Llais V2 could operate within these frameworks, making use of standard formats and ensuring that outputs could be easily reintroduced into editing timelines. This includes working with metadata such as EDL files and understanding the differences between platforms like Avid, Premiere and DaVinci Resolve.

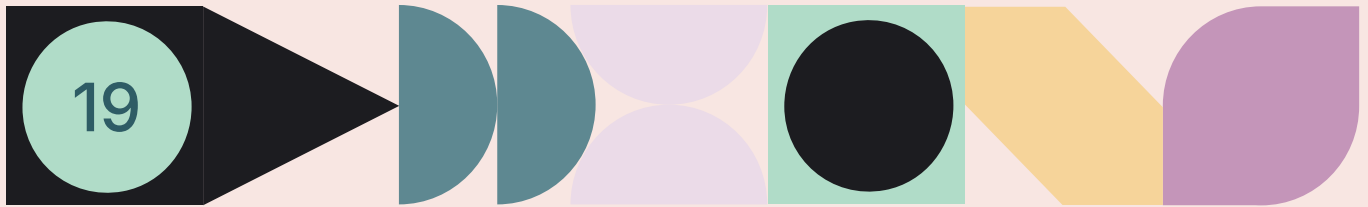
Encouragingly, there was a positive response to the idea of a system that could sit alongside these tools, particularly if it could be accessed through an interface and did not add unnecessary complexity.

The project analysed standard UK post-production environments (see appendix)

Key Findings

For adoption, Llais V2 must:

- Accept standard formats (e.g. EDL, XML)
- Return media with correct metadata
- Fit seamlessly into existing pipelines



testing with broadcast content

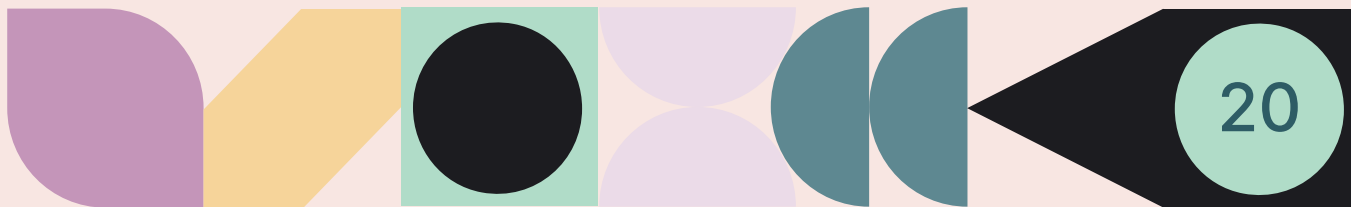
Testing the pipeline using real broadcast material was one of the most valuable aspects of the project. It provided a realistic view of how the technology performs under the kinds of conditions encountered in production.

S4C Archive:

Llais signed a contract with S4C allowing the project to access archive content. Any content to be used required clearance and authorisation from the person whose data would be trained and the production company.

The S4C legal team created a contract granting the Llais V2 project permission to use specified training data to develop and improve a Welsh-language AI model for translation and lip-sync purposes, subject to strict legal, security, and ethical conditions.

The data can only be used in isolated, non-production environments for the specific purposes, and cannot be uploaded to public datasets, reused for other AI projects, commercialised, or used to create outputs that reproduce or resemble the original data.



The agreement also prohibits the use of sensitive personal data and requires compliance with UK data protection laws, privacy-by-design principles, and recognized security standards.

The agreement gave S4C extensive oversight and enforcement rights, including the ability to audit the project, pause or revoke permission with short notice, and require deletion of all training data, models, and related artifacts in cases of misuse or termination.

The agreement was particularly valuable because it was available in Welsh, providing a useful reference point for Welsh-language AI terminology and legal phrasing. This helped support the development of consistent language and terminology for future Llais contracts, especially in emerging areas such as AI training, data governance, privacy, and machine learning compliance.

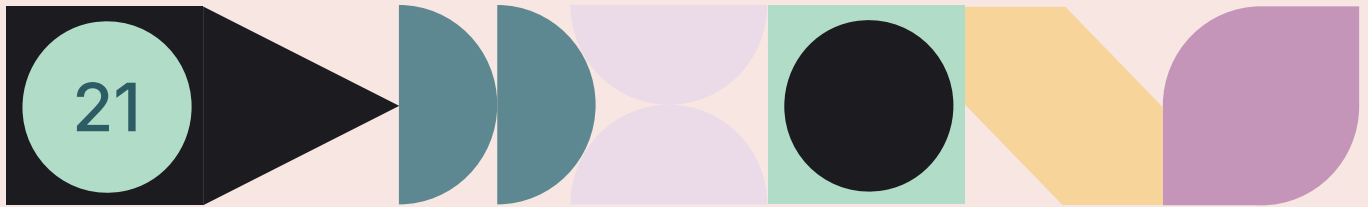
A number of programmes and genres were explored with the archive team. Due to the timeline of the Llais V2 project and the time required to clear the authorisation the team focussed on content where relationships with the presenters was already established.

Work on programmes such as *Ein Llwybrau Celtaidd* demonstrated that strong results can be achieved where the footage is of high quality and the subject is clearly visible. However, it also highlighted limitations. Lower resolution footage, challenging lighting conditions and rapid editing can all reduce the effectiveness of the lip-sync.

Similarly, work with archive material, *Gwanas I Gbara*, revealed additional challenges. Older footage often lacks the clarity needed for consistent results, and issues such as motion, shadow and occlusion can introduce visual artefacts. These findings suggest that there is additional processing and human intervention required when using archive footage.

Technical Insights

- Scene-specific training significantly improves results
- Audio alignment remains time-intensive
- Metadata (e.g. EDL files) is critical for workflow integration



Pilot Testing with Broadcast Content

Case Study 1:

Ein Llwybrau Celtaidd / Our Celtic Connections

Broadcast: S4C / TG4 – May 2026

Presenter: Ryland Teifi

Production company: Tinopolis

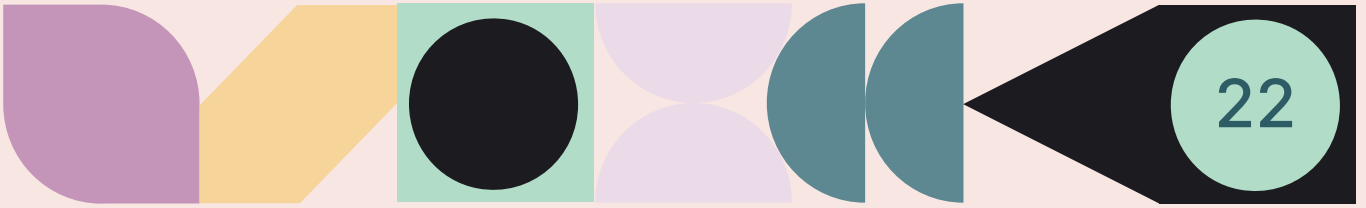
- Mixed results depending on footage quality
- High-quality close-up footage performed well
- Challenges:
 - Low resolution
 - Fast cuts
 - Lighting inconsistencies

Key Learning:

Scene-specific model training significantly improves output.

Examples: <https://www.ypod.cymru/about/llais-v2-rd-llwybrau-celtaidd/>





Case Study 2:

Gwanas I Gbara

Broadcast: S4C 2010

Presenter: Bethan Gwanas

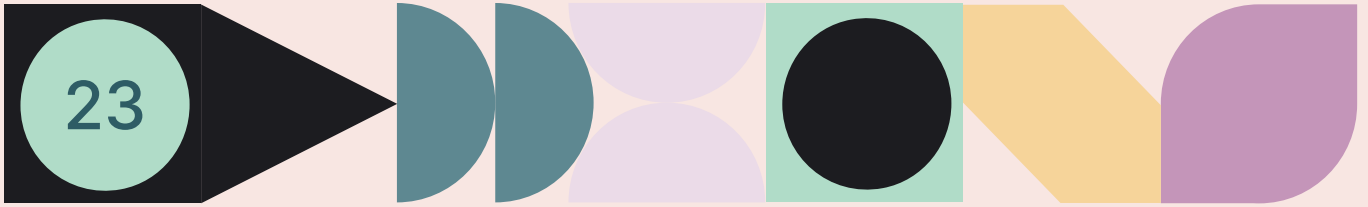
Production company: Telesgop

- Archive (2010) content posed challenges
- Issues:
 - Low resolution (SD footage)
 - Motion interference (hands, shadows)
 - Visual artefacts (flickering)

Key Learning:

- Older archive footage may require additional VFX processing
- AI-assisted rotoscoping (e.g. SAM models) shows potential but still requires manual QA

Examples: <https://www.ypod.cymru/about/llais-v2-rd-gwanas-i-gbara/>



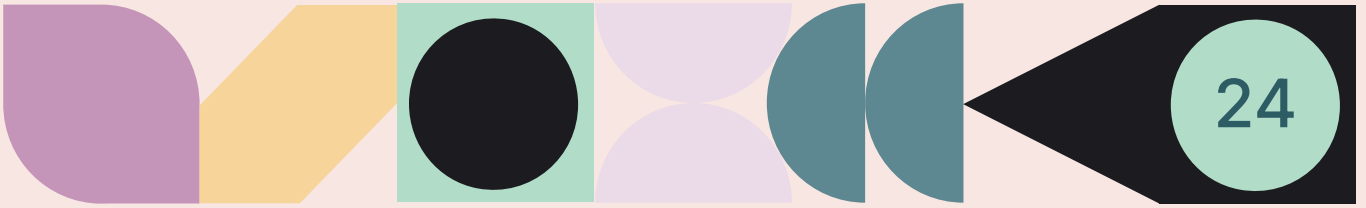
Broadcast Content Summary

Key Challenges

- Variability in footage quality
- Handling fast cuts and motion
- Audio alignment complexity
- Integration across different editing platforms
- Ethical and legal considerations
- Archive content limitations

Key Learnings

- Preparation of footage is critical
- Scene-specific training is more effective than general datasets
- AI cannot yet fully replace human oversight
- Industry adoption depends on seamless workflow integration
- Ethical frameworks must evolve alongside technology

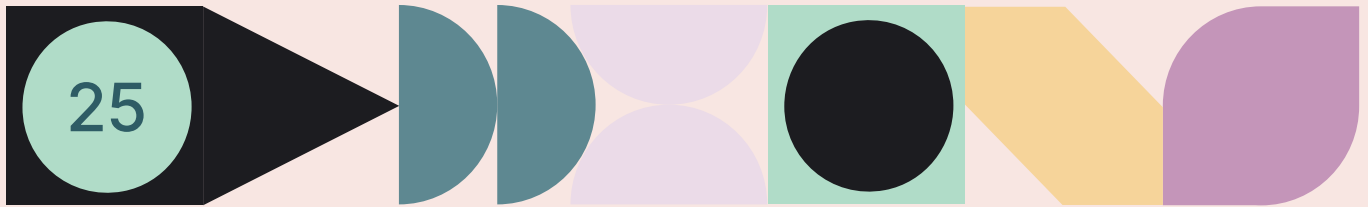


key challenges and learnings

Several key challenges emerged over the course of the project. Technical issues such as audio alignment and handling complex visual scenes are areas where the team can develop further. The quality of source material should be analysed as the system performs best under specific conditions.

The project reinforced a number of important lessons. It is clear that scene-specific training can significantly improve results, and that careful preparation of material is essential. It also became evident that, despite advances in automation, human oversight remains a crucial part of the process.

Beyond the technical, the importance of ethical considerations and industry alignment cannot be overstated. The success of this technology will depend not only on what it can do, but on how it is perceived and adopted by those working within the sector.



commercial potential & future development

Llais V2 has demonstrated that there is a credible commercial opportunity for AI-assisted localisation. The ability to adapt content for different audiences more efficiently has clear value, particularly for smaller language markets seeking to reach global viewers.

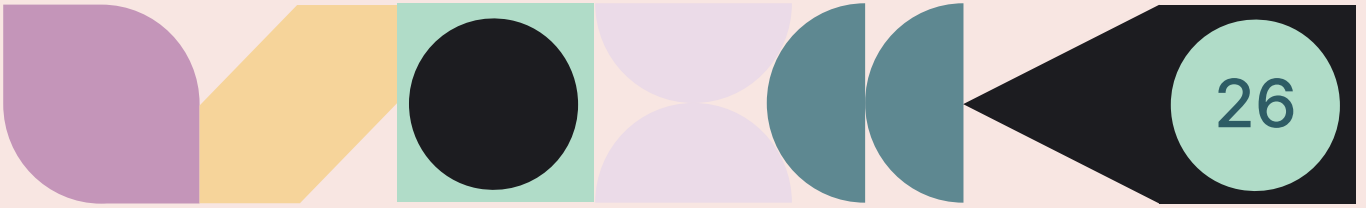
In addition to localisation, the technology offers potential benefits within post-production itself, providing new ways to refine and adapt content without the need for costly reshoots.

Looking ahead, further work is needed to improve automation, expand language capabilities and refine the system's performance across a wider range of content types. There is also an opportunity to explore new applications, including animation and other formats that may require tailored approaches.

The project has also begun to establish a clear pathway towards commercialisation. The team has been approached by two independent production companies to develop pilot projects using the technology, reflecting early industry interest and confidence in its potential. Through the course of the project, the team has developed practical experience and is well positioned to support real-world production workflows.

Initial exploration of a pricing model indicated that a fixed-cost approach is unlikely to be suitable. Instead, costs are expected to vary on a per-project basis, as they are dependent on factors such as the number of faces requiring training and the volume and complexity of voice inputs.

The next phase of development will focus on working with broadcasters and independent production companies to produce pilot content. These collaborations will be critical in testing the technology in live production environments, refining the commercial model, and demonstrating its value at scale.

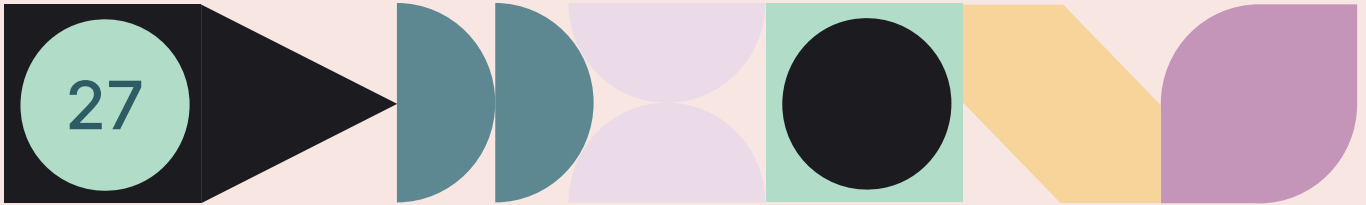


conclusion

Llais V2 represents a significant step forward in the development of AI-driven lip-syncing for media production. The project has moved beyond theory, demonstrating how the technology can be applied in practice and identifying the conditions under which it is most effective.

While challenges remain, the work has established a strong foundation for future development. Just as importantly, it has clarified the broader context in which the technology must operate, balancing innovation with ethical responsibility and industry needs.

Taken together, these findings point towards a future in which AI can play a meaningful and supportive role in media production, helping to extend the reach of content while preserving the creative contributions at its core.



[< Back to projects](#)

Project

S4C_Archive

Status: NEW

i

Generate Invite

Archive

Delete

Uploader Invite link

https://main.dj4uk91ljf52z.amplifyapp.com/upload/eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJwcm9qZWN0SWQIOiJmNjc2YjM4MS1iNjAwLTQ5YmQtYTZhYi1jYmRkMDM2ZWZhNWliLCJzZXNzaW9uSWQIOiIOM2VlNmI5OS1mNDlhLTQyNjYtOGVhZi02OGViZDFkODhhOWliLCJpYXQiOiE3ODAyNDY2NDEsmV4cCI6MTc4MDMzMzA0MX0uJ03L-J41rOQQScEx80iOBvTWy6LSXBVsTVlflNyLCgc

Expires: 01/06/2026, 17:57:21

Inventory

Refresh inventory

No assets yet. Run upload + ingest to populate inventory artifacts.

Admin Console

Projects

Refresh

Create projects and generate uploader links.

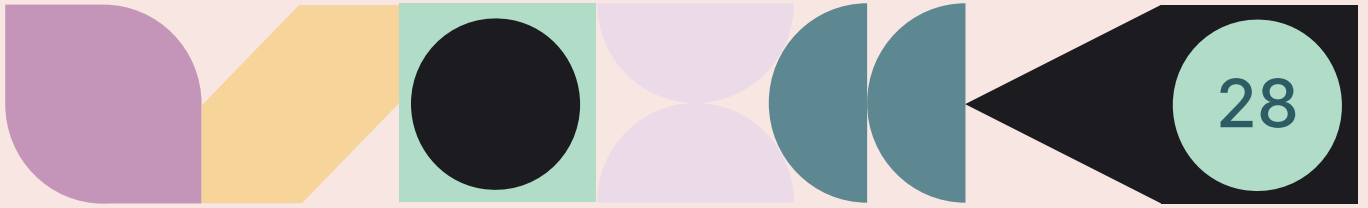
Create project

Project name

Create

Projects

Name	Status	Created	
S4C_Archive	NEW	31/05/2026, 17:56:21	Open
Rheiland	NEW	31/05/2026, 17:56:03	Open
new_test	NEW	23/02/2026, 12:04:58	Open



showcasing

Llais V2

A number of showcasing opportunities arose during the project timeline, in particular once examples of the content was available. Llais V2 was showcased at the following events.

Cymraeg y Cyfryngau Heddiw:

Friday 27 March 2026

Location:

Cardiff University

Presented Llais V2 at the Cymraeg y Cyfryngau Heddiw conference (the Welsh of today's media)

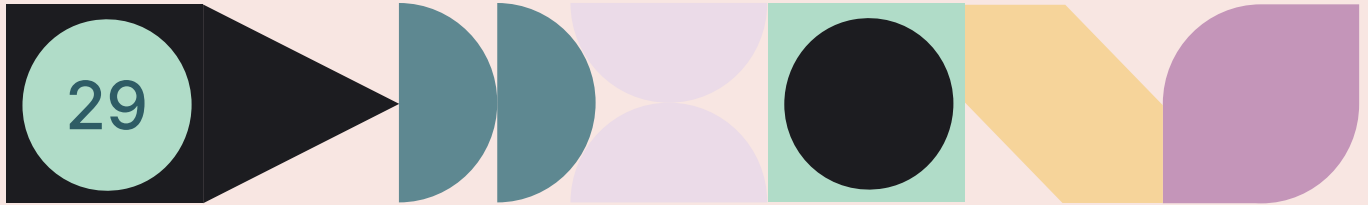
The conference was sponsored by S4C and the Coleg Cymraeg Cenedlaethol.

Cymraeg y Cyfryngau Heddiw

Bwriad y gynhadledd hon fydd
dwyn ynghyd academyddion
ac arweinwyr yn y diwydiannau
creadigol ac mewn sectorau
cysylltiedig er mwyn trafod
agweddau amrywiol ar
Gymraeg y cyfryngau heddiw.

27 Mawrth 2026, 09:30-16:00
Prifysgol Caerdydd





Plurilingmedia policy workshop:

27-28 April

Location:

University of Galway

Mixed language audiences in the context of changing consumption habits.

Presented Llais V2 as part of the Subtitling Audiovisual Minority Language Media



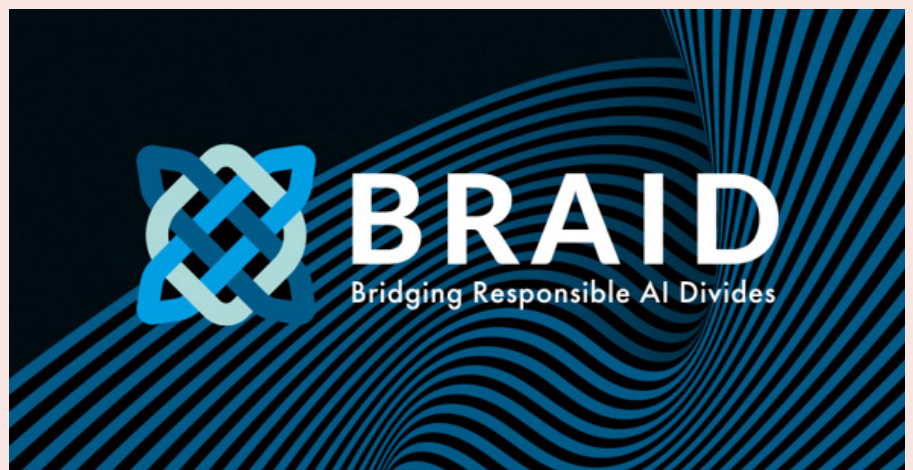
Llais is featured in the Braid Mooc – free short online course

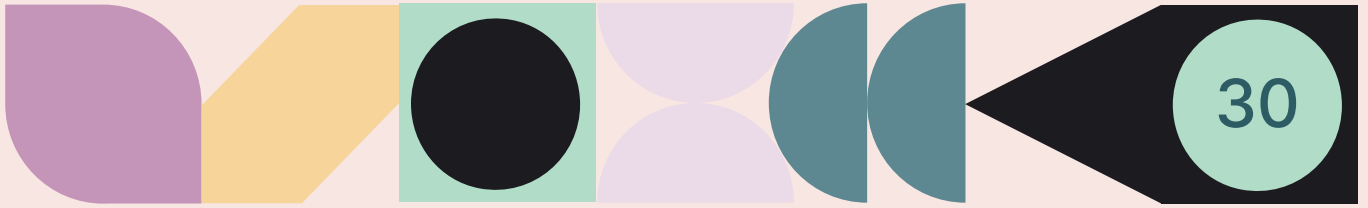
Responsible AI Innovation for SME's

The course aims to equip business leaders with enough knowledge to make informed decisions. The course is part of the Bridging Responsible AI Divides (Braid) research programme.

<https://media.ed.ac.uk/channel/>

An+Introduction+to+Responsible+AI+for+SMEs+%28BRAID%29/380203322





glossary

ADR (Automated Dialogue Replacement): The process of re-recording dialogue in post-production to improve audio quality, correct performance issues, or replace dialogue in another language.

AI (Artificial Intelligence): Computer systems designed to perform tasks that normally require human intelligence, such as speech recognition, translation, or image processing.

AI-Assisted Localisation: The use of artificial intelligence tools to adapt media content for different languages or audiences while supporting human creative workflows.

AI Lip-Syncing: Technology that modifies or generates lip movements in video so they match newly created or translated audio.

AI Voice Generation: The use of synthetic voice technology to create spoken dialogue using trained AI voice models.

Artefacts (Visual Artefacts): Unwanted visual distortions or errors created during image processing or AI generation, such as flickering, blurring, or unnatural movement.

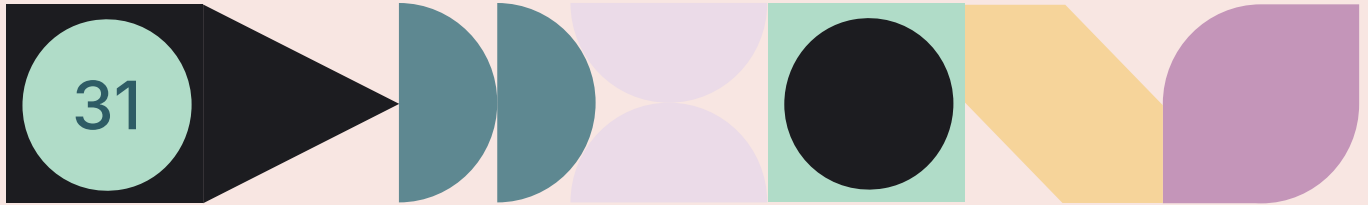
Audio Stems: Separate audio tracks containing isolated elements such as dialogue, music, or sound effects used during post-production.

Automation: The use of technology to complete repetitive or technical tasks with limited manual intervention.

Client-Facing Portal: An online interface that allows users to upload media, scripts, audio files, and other project materials into the workflow.

Data Governance: The management of how data is collected, stored, accessed, and used within a project or organisation.

Deepfake Technology: AI-generated media that realistically alters or recreates a person's appearance, voice, or actions.



Diarisation (Speaker Diarisation): An AI process that identifies and separates different speakers within an audio recording.

EDL (Edit Decision List): A text-based file used in video editing that contains information about clip order, timecodes, and edits within a timeline.

Ingest: The process of importing media files, scripts, and metadata into a production system.

Localisation: The adaptation of media content for different languages, cultures, or regions.

Machine Learning: A branch of AI where systems improve performance by learning patterns from training data.

Post-Production: The stage of media production that takes place after filming, including editing, sound design, visual effects, and finishing.

Resolution: The level of visual detail in a video image, typically measured in pixels.

Rotoscoping: A visual effects technique used to isolate or trace objects frame-by-frame within video footage.

SAM Models (Segment Anything Models): AI models used to automatically identify and separate objects or regions within an image or video frame.

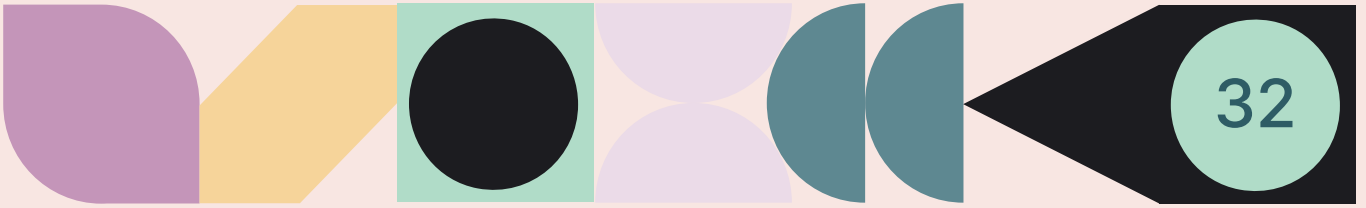
Scene-Specific Training: The process of training AI models using footage from a specific scene or performer to improve output quality.

Segmentation: The process of dividing media into smaller sections, often based on dialogue, scenes, or speakers.

Source Material: The original video, audio, scripts, or assets used within a production workflow.

Synthetic Media: Media content generated or modified using artificial intelligence technologies.

Translation Model: An AI system designed to translate text or speech from one language into another.



appendix

Summary of post production facilities in Wales

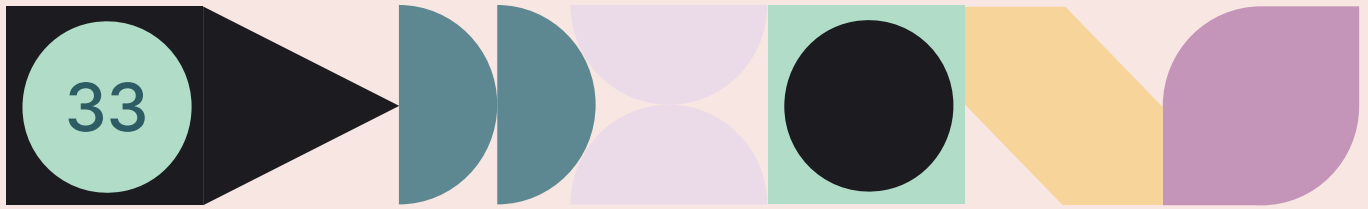
In the UK today, most professional post-production houses (film, TV, commercials, broadcast and high-end online work) tend to use a set of well-established industry-standard software tools - each for specific parts of the workflow. These reflect what major facilities equip their suites with and what studios and broadcasters expect on delivery.

Editing

- **Avid Media Composer:** Still considered the industry standard for long-form and broadcast editing because of its robust media management and team collaboration features. It's widely used by TV editors and bigger post houses.
- **Adobe Premiere Pro:** Very common especially for commercials, corporate projects, promos and increasingly in feature/TV workflows; integrates well with After Effects for graphics.
- **DaVinci Resolve:** Originally a colour grading tool, now widely used for editing too, especially where the project also needs integrated grading and finishing.

Colour Grading & Finishing

- **DaVinci Resolve Studio:** A dominant tool for professional colour grading and increasingly used end-to-end (edit → grade → delivery).
- **Baselight / Baselight Editions:** High-end grading systems used in feature film and high-end TV finishing (found in some UK houses).



VFX & Compositing

- **Foundry Nuke:** A core industry tool for VFX compositing in film and broadcast. (Founded in the UK, widely used globally.)
- **Adobe After Effects:** Often used for motion graphics, compositing, tracking and smaller VFX tasks.
- **Autodesk Flame / Smoke:** Used in higher-end finishing and effects workflows.

Audio Post & Mixing

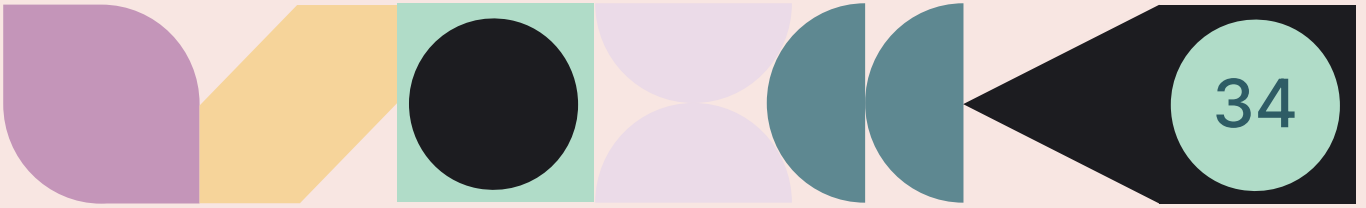
- **Avid Pro Tools:** The dominant standard for sound design, mixing and broadcast-ready audio delivery.

Why these matter

- **Broadcast compatibility:** Major broadcasters and facilities expect edits and deliverables that work with shared storage and conform systems — and Avid workflows still lead here.
- **Integrated colour and finishing:** Tools like Resolve let teams edit and grade in one environment.
- **Graphics & VFX workflows:** After Effects and Nuke cover everything from motion graphics to complex compositing.

Trends

- **Resolve is growing** into more editing roles because of its all-in-one workflow.
- **Premiere is strong** in advertising and cross-platform projects.
- **Final Cut Pro** is less common in professional UK post houses compared to these tools (more popular with solo editors and corporate work), though you'll see it occasionally.



EDL Files

The Llais V2 project could use an EDL file that is used across a range of platforms

What an EDL contains

An EDL is a plain-text file that lists:

- Which **source clips** were used
- **In and out timecodes** for each clip
- The **order** of the clips on the timeline
- Basic **cuts and transitions** (usually just straight cuts)

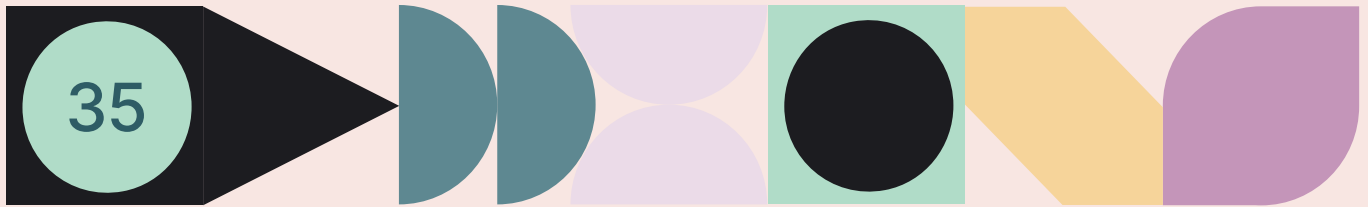
It basically says: "Take clip A from 01:02:10:05 to 01:02:18:12, then cut to clip B..."

How this works for Llais V2

Ideally Llais V2 should be receiving material from an Edit to dub and lipsync.

Llais needs to balance the volume of material we get so we primarily use the AI Model the footage that will be in the programme, but also get raw material for scenes for training and get separated audio stems per actor when available.

Delivering content back with the right meta data, so that the content can be slotted straight back in, which will be easier if the post production facility is working on Premier / DaVinci, but more difficult if using Avid.



Technical R&D Summary

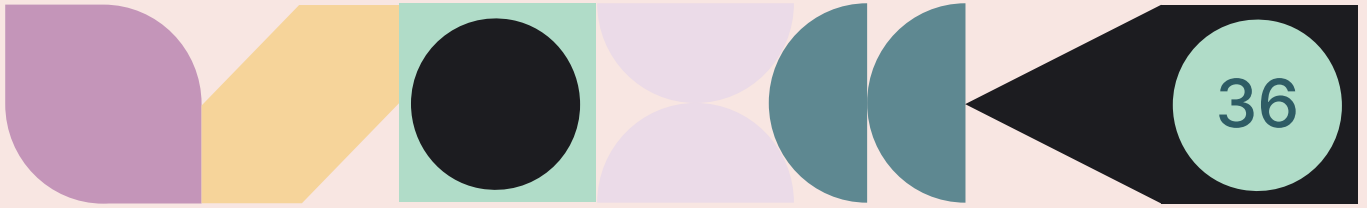
Alongside the continued refinement of the Lipsync AI pipeline developed during **Llais V1**, we analysed the main pain points in the production workflow, with a particular focus on opportunities for time savings and quality improvements.

The first challenge we identified was the amount of time spent on transcription, translation, and, crucially, aligning translated audio to video. In many cases, translated dialogue had to reach the alignment stage before it became clear that it did not fit the available timing and required a rewrite. This created inefficiency because translators or writers had to be brought back into the process late, after technical alignment work had already begun.

We therefore needed to automate this process in a way that supported earlier human intervention. The goal was not simply to replace manual rewriting, but to make it easier for translators, writers, and editors to identify timing problems while the script was still being developed. This led to the creation of an MVP version of **DubTool**.

During the development of DubTool, it became clear that the quality of AI transcription, particularly for Welsh, was a significant limitation. This led to a sub-project exploring the relative quality of different transcription tools and investigating how Welsh transcription could be improved.

The third major challenge was the time spent sourcing appropriate training material for the AI lipsync process, and then identifying and cutting suitable clips. During trials, production partners were usually able to provide the programme or clip requiring lipsync dubbing quickly, since final programmes are readily accessible. However, we often had to wait months for the raw footage behind the finished programme. This led us to investigate how we could make delivery easier for clients while also giving us faster access to the assets needed for training and production. This work culminated in our final prototype: a client-facing **Ingest Portal**.



DubTool

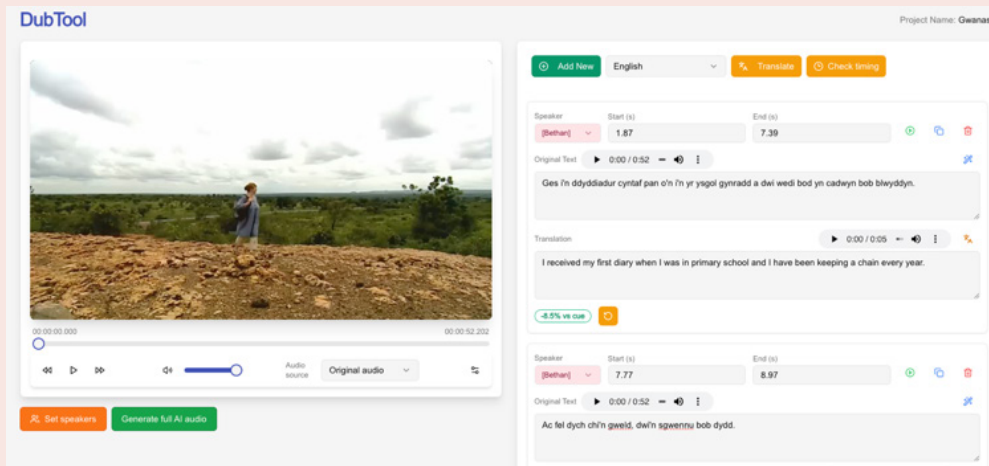
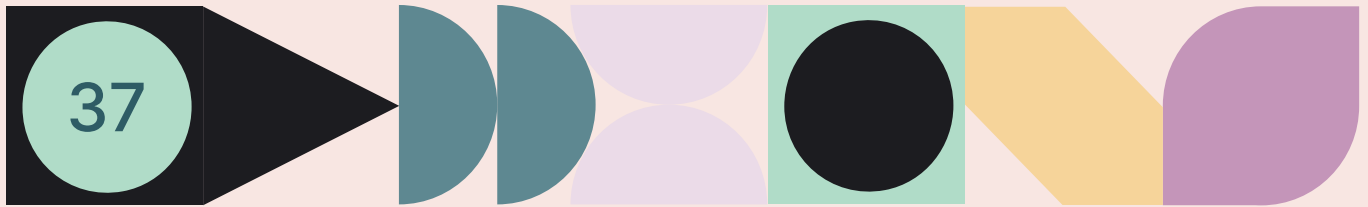
DubTool is a prototype web tool for preparing dubbed or translated media. It gives production teams a single place to turn source media and subtitles into structured localisation assets.

The tool helps editors review timing, speaker attribution, and subtitle content alongside the media player. It supports faster translation workflows through bulk subtitle translation, while still allowing straightforward manual subtitle editing. It also prepares exportable outputs for downstream dubbing, review, and localisation workflows.

The basic workflow is:

- Upload a source video.
- Automatically create a timecoded transcription/subtitle file.
- Edit subtitle entries while watching the video.
- Assign speakers, enabling separate scripts for different voice actors.
- Translate subtitle content and allow easy correction.
- Automatically test the spoken duration of translations using AI voices.
- Support iterative edits where translated lines do not fit the available timing.
- Export dub-ready, timecoded scripts per voice.

A screenshot of the DubTool web interface, specifically the 'Select or Add a Dubbing Project' dialog box. The interface is clean and modern, with a white background and dark text. At the top, there's a title bar with a close button (X). Below the title, there's a 'Project name' field with the text 'DubTool Project'. Underneath, there's a section for 'Choose your working languages' with 'INPUT' and 'OUTPUT' dropdowns. The 'INPUT' dropdown is set to 'Auto-detect' and the 'OUTPUT' dropdown is set to 'English'. There are two checkboxes: 'Code-switching (between input language & English)' which is unchecked, and 'Include "umm..." disfluencies' which is checked. Below this, there's a section for 'Select transcription/diarisation provider' with two radio buttons: 'Assembly AI' (selected) and 'Speechmatics'. The 'Assembly AI' option has a subtext 'Lower-cost and better diarisation.' and the 'Speechmatics' option has a subtext 'May be better for less common languages.' At the bottom, there are two 'Upload media for processing' fields, each with a 'Choose file' button and 'No file chosen' text. Below these, there's a 'Clean audio (optional)' field with a 'Choose file' button and 'No file chosen' text. At the very bottom, there are three large buttons: 'Upload', 'Diarise', and 'Go to Editor'. A small note at the bottom right says 'You must upload and diarize media before proceeding.'



Key Features

Media upload and playback: source media can be uploaded and previewed in-browser while subtitles are edited.

Subtitle editor: users can add, edit, and manage subtitle entries against the media timeline.

Speaker management: subtitle lines can be assigned to individual speakers, supporting clearer dubbing workflows and separate voice scripts.

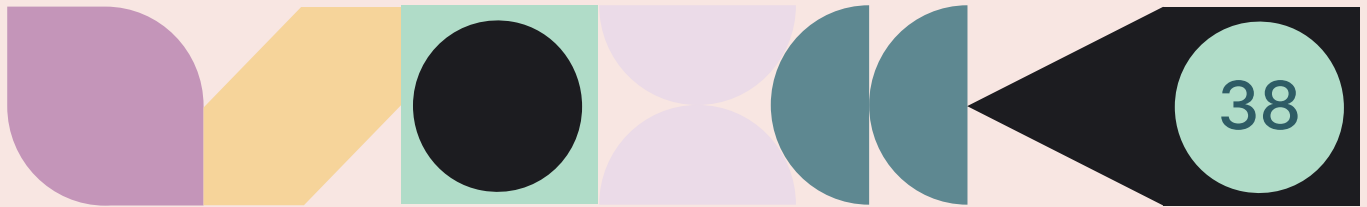
Bulk translation: subtitle content can be translated as part of the localisation process.

Duration testing: translated lines can be tested using AI-generated voices to check whether the spoken duration fits the available timing.

Review workflow: editors can inspect subtitle timing, speaker allocation, and translated content alongside the media player before export.

Export flow: edited and translated subtitles can be exported as dub-ready scripts or packages for downstream production.

DubTool was used in our most recent test and showed promise in speeding up the preparation and translation workflow. However, the test also highlighted the need to improve Welsh transcription, which was not as well supported by the AI transcription products we had been using, including AssemblyAI and Speechmatics.



Welsh Transcription Trial and Research

A screenshot of a web interface for uploading audio and getting Whisper Welsh SRT. The interface has a white background with a light blue border. At the top, it says "Upload audio & get Whisper Welsh SRT". Below this, there is a "Choose file" button and a "No file chosen" text. To the right, there is a blue "Upload & Transcribe" button. Below the upload area, it says "Idle." and "Transcript status and messages will appear here." Below this, it says "Completed downloads (latest first)". At the bottom, there is a section titled "Preview: Speechmatics vs Whisper" with a "Waiting for completed job..." status. The preview section has a table with columns for "Timecodes", "Speechmatics", "AssemblyAI", and "Whisper". The "Timecodes" column shows "No preview yet."

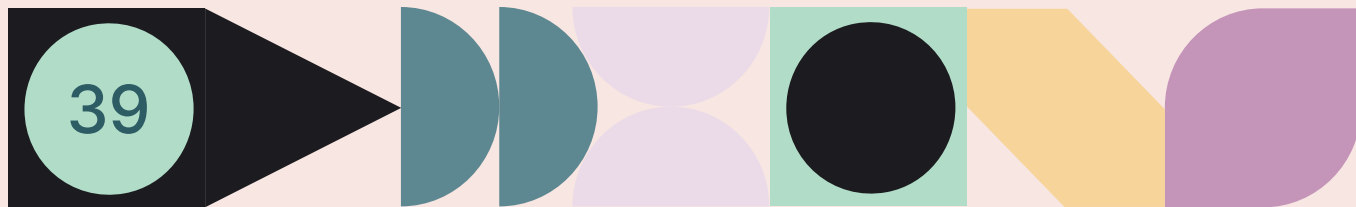
To investigate the transcription issue, we used a comparison tool to test the accuracy of the different transcription systems being used in DubTool. These included **Speechmatics**, **AssemblyAI**, and a locally installed **Whisper variant trained by Canolfan Bedwyr**. The tool allowed direct comparison of the outputs from each system.

The results showed that the Whisper-based tool was generally more accurate most of the time. However, it also had specific weaknesses that the other models did not share. For example, in our tests, Speechmatics was generally the least accurate overall, but performed better on place names than the Whisper model. This may have been because the Whisper model was trained on material where place names had been removed for privacy reasons.

These findings led us to consider whether the best approach would be to use all three transcription models in combination, and then use a **Large Language Model (LLM)** to assess which output was most likely to be correct for each phrase.

To test this idea, we set an MSc dissertation topic for a student on the Data Science course at Cardiff University. The student, Daniel Hiscox, carried out a detailed piece of research statistically testing the accuracy of the individual transcription models and the LLM-assisted approach. His work supported the hypothesis that an LLM could make a meaningful improvement to transcription quality by selecting or mediating between model outputs.

We plan to implement this approach within DubTool in future development.



Production Asset Delivery and Training Material

In addition to improving transcription quality, we also found that we need to ensure the correct range of production assets and documentation is available earlier in the workflow. In particular, wider source material is required to train the AI model before dubbing each individual speaker.

This material can come from outtakes or from other sections of the production. However, for some projects this was challenging because there was little or no spare footage outside the specific performance to be dubbed. Ideally, productions need to be encouraged to deliver this material from the original project folders or editorial bins, where it may already exist but is not usually included in standard delivery.

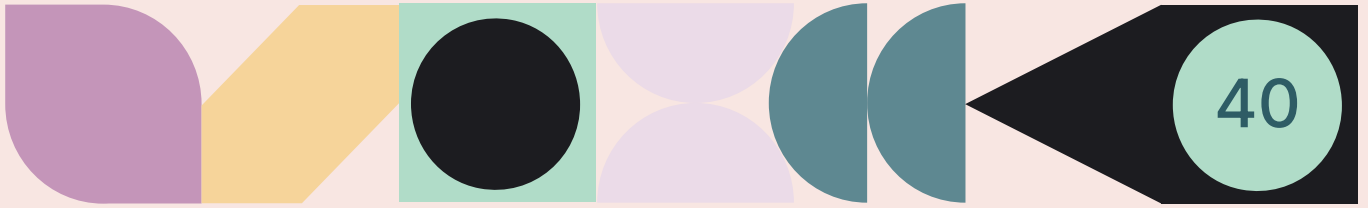
Alongside this delivery challenge, we also needed to clarify where the material being dubbed sits within the production pipeline. For example, there is a significant difference between dubbing a final graded programme and dubbing material before final edit or assembly.

Dubbing specific shots before assembly has several potential advantages. It gives editors more flexibility to lengthen or shorten shots to allow for differences in dialogue duration between languages. It also reduces the likelihood that the lipsync system has to infer mouth positions when a cut occurs mid-dialogue. In addition, applying final grading after lipsync may help disguise any visual imperfections.

However, working shot by shot is more complex than processing a whole finished programme. It requires receiving, editing, tracking, and returning a much larger number of files. It can also create practical issues. For example, if captions, name straps, or other on-screen graphics have been added later in the post-production process, they may not appear on the material returned after lipsync processing.

After discussions with one of the TV production companies about their post-production chain, it became clear that we needed a tool that would make delivery easier for the client while also tracking each shot and returning the lipsynced version in a format compatible with their existing workflow.

Based on further research and discussions with the production company, we decided to build a prototype ingest portal that would work with export tools already available to production teams. The initial focus was on delivery from **Adobe Premiere**, with the potential to expand to import/export workflows for **DaVinci Resolve** and **Avid**.



Llais Lipsync Ingest Portal

The **Llais Lipsync Ingest Portal** is a web portal for ingesting editorial delivery folders, validating package contents, parsing OTIO timelines, generating inventory artifacts, and previewing media for admin review.

App Purpose

The portal gives administrators a single place to create projects, issue secure uploader links, and monitor ingest outcomes.

It gives uploaders a guided flow to preflight and upload large folders using resumable multipart upload.

It produces machine-readable ingest outputs, including `assets.json`, `clip_usage_index.json`, warnings, job reports, and preview status, for downstream workflows.

Key Features

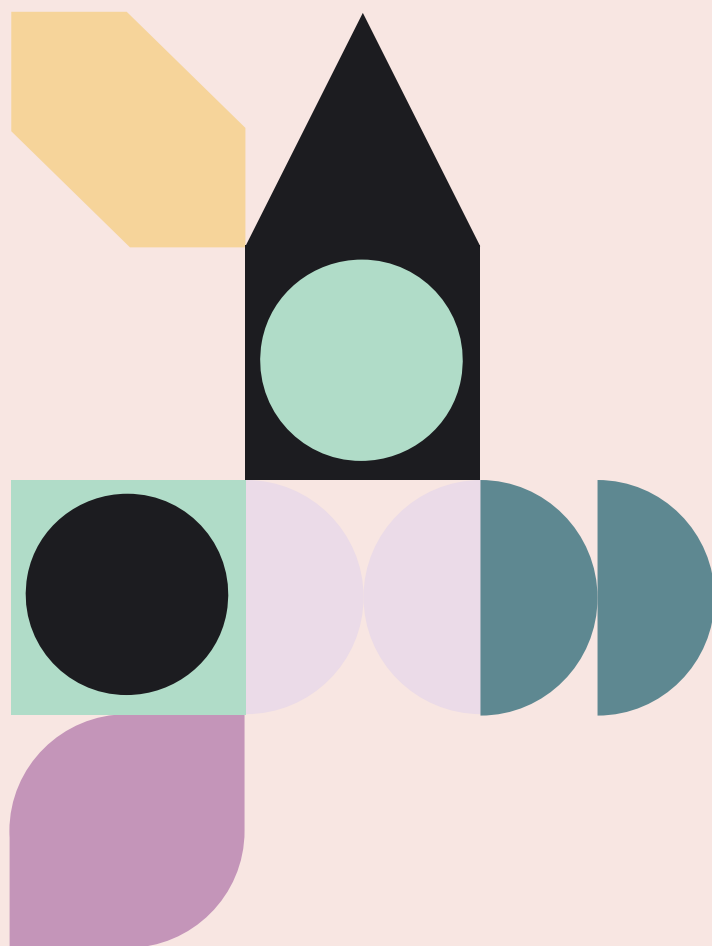
Admin project management: create, list, and view projects; generate invite links; archive and delete projects.

Uploader portal: preflight checks, multipart upload endpoints, and finalize marker trigger.

Ingest pipeline: validates delivery packages, parses OTIO timelines, resolves media references, and sets preview media.

Proxy pipeline: uses MediaConvert when no `Reference/roughcut_ref.mp4` file exists.

Inventory and preview UI: provides artifact downloads and in-browser playback, with MP4 and HLS support.



Thanks for reading!

2026

media
cymru × pdr.