

ATISN 27572 follow up to ATISN 27499
(Welsh Government collaboration with
Cosine)

Meetings

“RE: Intros – Cosine / Welsh Gov”

- 15/07/26
- Teams meeting
- [WG Official 1], [Cosine 1]
- Introductory call between the Welsh Government and Cosine.

“Cosine / WG – Welsh language”

- 20/07/26
- Hybrid meeting
- [WG Official 1], [WG Official 2], [WG Official 3], [Cosine 1], [Cosine 2]
- Follow up to the introductory call to discuss the Welsh language in particular.

“RE: Cosine / WG Welsh language”

- 07/09/26
- Teams meeting
- [WG Official 1], [WG Official 2], [WG Official 3], [Cosine 1]
- Follow up to “Cosine / WG – Welsh language”.

“Cyd-destun Cosine”

- 22/09/26
- Teams meeting
- WG [WG Official 1], [WG Official 2], [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]
- Background to clearance of ATISN 27499 response.

Emails

ATISN 27572 – Email – 20260715 – “Re: Welsh Language LLM”

From: [WG Official 1]

Sent: 16 July 2026 09:29

To: [Cosine 1]; [WG Official 2]; [WG Official 3]

Cc: [Cosine 2]

Subject: ATB/RE: Welsh Language LLM

Hi both, would Monday suit? Free 11am or 1pm.

Adding [WG Official 3] who leads on Welsh language use.

[WG Official 1]

Oddi wrth: [Cosine 1]

Anfonwyd: Dydd Mercher, 15 Gorffennaf 2026 17:55

At: [WG Official 1]

Copi/Cc: [WG Official 2]; [Cosine 2]

Pwnc: Re: Welsh Language LLM

Thanks [WG Official 1]. Copying [Cosine 2] in our ML team who would love a conversation here. [WG Official 2] - let us know when you have time over the next few days for a call.

[WG Official 1] I'll send a separate email re. our other actions.

[Cosine 1]

On Wed, 15 Jul 2026 at 17:53, <[WG Official 1]> wrote:

Hi [Cosine 1]

Great catching up today, thank you very much for your time. Further to our conversation relating to Welsh Language support, can I please introduce you to [WG Official 2], who has been leading on the development of Welsh Language AI in recent months, and has developed several data assets for the training of LLM's.

Please let us know your thoughts on next steps.

Kind regards

[WG Official 1]

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

ATISN 27572 - Email - 20260720 - "RE: Cosine & Welsh Gvt"

On Mon, 20 Jul 2026 at 21:30, [WG Official 2] wrote:

Thanks [Cosine 2], and it was probably mine also, certainly the most exciting one. And apologies for the delay in getting back to you, I was hit bug end of last week, and only just catching up with emails.

We had a great call with [Cosine 3] today and the Welsh language team have already shared a series of files ***[Explanatory note: “files” refers to the links to publicly available resources as shared previously under FOI ATISN 27499]*** to get things moving.

[redacted information]

Thanks again to you both, exciting times.

KR

[WG Official 2]

From: [Cosine 2]
Sent: 15 July 2026 18:55
To: [WG Official 2]
Cc: [Cosine 1]
Subject: Cosine & Welsh Gvt

Hi [WG Official 2],

Great to speak to you today. It was the best meeting of my day (and I promise I don't always say that!)

[redacted information]

- Thank you for the intro to [WG Official 3]. We will take forward

Keep in touch,

[Cosine 2]

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â

Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

ATISN 27572 - Email - 20260730b - "Re: Potential Cosine partnership announcement"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Iau, 30 Gorffennaf 2026 20:06

At: [Cosine 1]; [WG Official 2]

Copi/Cc: [Cosine 2]; [Cosine 3]

Pwnc: RE: Potential Cosine partnership announcement

Hi [Cosine 1]

Thanks for getting in touch. Let us look into this internally and we will come back to you to confirm Monday/Tuesday next week.

Kind regards

[WG Official 1]

From: [Cosine 1]
Sent: 30 July 2026 07:19
To: [WG Official 1]; [WG Official 2]
Cc: [Cosine 2]; [Cosine 3]
Subject: Potential Cosine partnership announcement

Hi [WG Official 1], [WG Official 2],

Hope all well! Quick one - which I don't believe we've discussed. Would you be happy for us to work on a joint press release about our engagement with you embedding Welsh language data into our Lumen model training? We'd love to publicise how we're working together.

If this is premature - or you'd rather not for any reason - then absolutely fine.

Best,

[Cosine 1]

ATISN 27572 - Email - 20260721 - "RE: Welsh open data"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mawrth, 21 Gorffennaf 2026 15:06

At: [Cosine 1] ; [WG Official 2]

Copi/Cc: [WG Official 3] ; [Cosine 2]

Pwnc: RE: Welsh open data

Thanks [Cosine 1] .

From: [Cosine 1]

Sent: 21 July 2026 14:53

To: [WG Official 1]

Cc: [WG Official 2] ; [WG Official 3] ; [Cosine 2]

Subject: Re: Welsh open data

And just to say: we've integrated a lot of this data now so thank you very much!

On Tue, 21 Jul 2026 at 14:51, [Cosine 1] wrote:

Hey all,

Just to follow up on this. I did a big audit of these sources and how they can fit into our training data mix yesterday.

I thought I'd share some of that information with you as it could be useful feedback at least from the point of view of language model training. Feel free to ignore if it isn't or if it's already obvious!

A disclaimer: GPT 5.6 Sol performed the full audit and produced the attached table (otherwise I simply would not have had the time). However, I directed it to review the sources specifically in relation to our training pipeline so hopefully there's some useful context on what we care about.

That being said, it blindly audited a few rows that evidently weren't intended to be data sources, and misunderstood the context in some cases and I'm not sure if I completely trust the stats on dataset size, so do take this with a pinch of salt.

Here is some very rough context on the typical stages of pretraining for LLMs, in case you don't already have it:

1. Pretraining - high-volume, diverse, varied and unstructured data, potentially containing lower-quality sources. This uses the standard next-token prediction objective: it's just about modelling stuff on the internet.
2. Mid-training or continued pretraining (CPT) - more specialist data sources aiming to increase knowledge in certain target areas. Can also start increasing the length of documents that the model supports as input here.
3. Instruct tuning - this essentially turns the model into a chatbot. It's a bit less about knowledge and more about human interaction as opposed to blindly regurgitating text it's seen during pretraining.
4. Reinforcement learning - a different learning objective to really push performance in areas such as coding, synthesis of information, search and alignment/safety.

At Cosine, we're focusing on stages 2-4. Increased Welsh representation can probably make the most impact in stage 2 and potentially stage 3.

To synthesise some of the audit findings in my own words:

- the short-sequence datasets consisting of floating paired English/Welsh sentences/utterances could make most sense in stage 3 where they can be framed as translation tasks but they aren't as useful in stage 2
- the audit's preoccupation with document/session IDs is because with these, we can concatenate sentences together into longer documents. Long-form data (anything from 1k to 700k words) is generally really useful to us because it helps the model deal with the longer text sources it's likely to see in real life. It can also really help with things like register because it can help the model to learn long range consistency in tone. Any more data of this kind would be really valuable to push performance in Welsh language.
- there were a few mentions of licences not being as permissive as we need for training, but I'm sure that's on your radar already

Some other points that feel relevant:

- Evaluation benchmarks are also generally really useful for us to track progress. I haven't done much research into these at all but [this](#) looks promising. Any further pointers are always appreciated!
- Generally speaking, the more these data sources can be collated and uploaded to huggingface in a standard format, the more likely they're going to be included in LLM training. Looks like this is already happening in some places and the techiaith group you linked look like they're making loads of progress with that [here](#) which is great.

I'm not sure if any of this is useful, but hopefully it gives you a general idea of the kind of data that is most valuable for us (and generally anyone training large language models).

All the best,

[Cosine 1]

On Mon, 20 Jul 2026 at 14:08, [Cosine 1] wrote:

Hi [WG Official 1] ,

Great to meet you all too! Excited to see what we can do here.

Thanks very much for all these links. I'll get working to integrate these into our pretraining corpus. Hopefully from there we can get a better idea of the volume and any potential gaps.

We're definitely open to the HeloBlod data too if it's straightforward to obtain.

Yes, an ongoing process for any new resources you find would be great. Maybe one to discuss at our catch up in September but from our point of view, a simple location/feed/page/email with a digest of resources as they're added would probably be easiest.

I've also already started looking a little into this problem of register and as I'm sure you already know, it largely seems to be caused by data scarcity for more colloquial/spoken Welsh. It sounds like you're already addressing this so that's promising. There's potential in future to integrate these aspects into our model evaluation pipeline too so that we can quantify performance there.

All the best,

[Cosine 1]

On Mon, 20 Jul 2026 at 12:09, [WG Official 1] wrote:

Hi [Cosine 1],

Great to meet you this morning.

The whole of the Welsh Government website is licensed under Open Government License (CCBY). This includes translation memories, giving you access to aligned sentences in tmx files. Other data are collected through Welsh Government funded academic research and supported crowdsourcing. Welsh Government also fund volunteer crowdsourcing campaigns with OpenStreetMap and Wikipedia. E.g...

- Transcriber, <https://trawsgrifiwr.techiaith.cymru/>
- Common Voice CY, <https://commonvoice.mozilla.org/cy/datasets>
- Paldaruo, <https://techiaith.cymru/data/corpora/paldaruo/?lang=en>
- Tagged sentences, <https://github.com/techiaith/brawddegau-tagiedig>
- DigiGrid, <https://digigrid.cymru/>
- CorCenCC, <https://corpus.corcenc.org/?language=en>
- Corpus of sentences <https://techiaith.cymru/data/corpora/brawddegau/>
- Podcast audio with verbatim transcriptions, <https://git.techiaith.bangor.ac.uk/data-porth-technolegau-iaith/banc-trawsgrifiadau-bangor>

- Bilingual word embeddings, <https://github.com/cardiffnlp/en-cy-bilingual-embeddings>
- BydTermCymru, tmx files, Translation memories | GOV.WALES
- Bangor University translation memories, https://cofion.techiaith.cymru/cy/search#cy_niwral
- <https://techiaith.cymru/cwmwl/api/lemmateiddiwr/>
- Termiadur addysg Bangor, <https://www.termiaduraddysg.cymru/?lang=en>
- Collection of URLs containing Welsh text, https://github.com/techiaith/urls_cymraeg
- Legislative data in Welsh/English, <https://www.legislation.gov.uk/cy>

We also have at least a million words of similarly quality-controlled bilingual text from our HeloBlod service we can look to share with you as well. This all includes a significant amount of thematic variation.

We can also continue to look for other resources as well, more technical beyond corpora, and a way of creating an ongoing sharing process with you – to discuss?

Diolch,

[WG Official 1]

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

ATISN 27572 – Email – 20260813 – “YML/FW: ATB/RE: Potential Cosine partnership announcement”

From: [WG Official 1]

Sent: 13 August 2026 12:25

To: [WG Official 2]

Subject: YML/FW: ATB/RE: Potential Cosine partnership announcement

Cosine wedi danfon hwn draw.. Beth yw'r ffordd gorau i wneud rhywbeth ar y cyd? Un PR joint neu fel nethon ni ar gyfer [ChatGPT learns Welsh | GOV.WALES](#), so un yr un efo exchanged quotes?

[WG Official 1]

ATISN 27572 – Email – 202609825 – “Cyhoeddiadau tech”

From: [WG Official 1] <[WG Official 1]>

Sent: 25 August 2026 14:30

To: [WG Official 2] <[WG Official 2]>

Subject: Cyhoeddiadau tech

Hi [WG Official 2],

Byddai'n briffio CMEWL ar y 9fed o Fedi ar Technoleg ac AI Cymraeg - briff CMEWL.docx, gan obeithio bwrw mlaen gyda'r cyhoeddiadau tech o w/c 14 Medi ymlaen.

Ar gyfer y bartneriaeth Cosine, ni'n gweithio ar hwn ar y cyd gyda Swyddfa AI felly nai goïo ti fewn i'r thread yna ar wahân.

Yn y cyfamser, use case newydd ar Copilot Cymraeg newydd fynd lan ar ein tudalen tech, Technoleg Cymraeg a deallusrwydd artifisial (AI): diweddariadau | LLYW.CYMRU

Diolch eto,

[WG Official 1]

ATISN 27572 - Email - 20260921 - "RE: For review by 12 noon Thursday 24 September please"

From: [WG Official 1]

Sent: Monday, 21 September 2026 15:02

To: [WG Official 2]; [WG Official 3]; [WG Official 4]

Subject: RE: For review by 12 noon Thursday 24 September please

Thanks [WG Official 2],

All look good to me.

Diolch yn fawr iawn am arwain ar hyn

[WG Official 1]

From: [WG Official 1]

Sent: 21 September 2026 15:52

To: [WG Official 2]; [WG Official 3]; [WG Official 4]

Subject: For review by 12 noon Thursday 24 September please

Hi bawb,

Could you please review the following draft responses? Once you've confirmed you're happy with the wording, we'll respond from Cymraeg 2050.

* Two WQs (by 12pm Thursday 24th Sept)

WQ100618(e) - John Clark - data Cosine.docx

WQ100619(e) - John Clark - partneriaeth Cosine.docx

* FOI

(attached iShare link)

* Welsh language Commissioner complaint

CS1551 Cwyn Safonau\u0027r Gymraeg.docx

* Combined response to two TOs

(attached iShare link)

Diolch,

ATISN 27572 - Email - 20260907c - "RE: For input today - PN On Cosine partnership"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 7 Medi 2026 12:05

At: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Copi/Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7];

[WG Official 8]; [WG Official 9]

Pwnc: RE: For input today - PN On Cosine partnership

Thanks [WG Official 2] looks good to me – thanks!

From: [WG Official 1]

Sent: 07 September 2026 12:03

To: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Subject: RE: For input today - PN On Cosine partnership

Sorry [WG Official 3][WG Official 2] and I conversing in the background on this one. We have an updated suggestion:

“AI has the potential to bring real benefits to our economy, public services and communities. By working with sovereign AI companies like Cosine, we can help build the skills, knowledge and capability Wales needs to make the most of these technologies, while supporting innovation that reflects our national priorities.

This will help strengthen Wales’ long-term position in AI, and is especially important for the Welsh language, helping to ensure new technologies work for people in the language of their choice.”

[WG Official 1]

ATISN 27572 - Email - 20260907b - "RE: For input today - PN On Cosine partnership"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 7 Medi 2026 11:58

At: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Copi/Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Pwnc: RE: For input today - PN On Cosine partnership

Looks good. Thank you.

From: [WG Official 1]

Sent: 07 September 2026 11:51

To: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Subject: RE: For input today - PN On Cosine partnership

Thanks [WG Official 2] – how about the below?

Cabinet Minister for Enterprise, Connectivity and Energy, Adam Price, said: “AI has the potential to be of great benefit to the economy and many areas of public life. By working in partnership with strategically important companies like Cosine, who are a UK Sovereign AI company, we can make sure those opportunities are realised and that people in Wales can use the technology in their preferred language.”

[WG Official 1]

ATISN 27572 - Email - 20260907a - "RE: For input today - PN On Cosine partnership"

Oddi wrth: [WG Official 1] <[WG Official 1]>

Anfonwyd: Dydd Llun, 7 Medi 2026 11:11

At: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Copi/Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7] <[WG Official 7]>; [WG Official 8]; [WG Official 9]

Pwnc: RE: For input today - PN On Cosine partnership

Thanks [WG Official 2],

Can we add some words about the fact Cosine is a UK Sovereign AI company and strategically important as well as important from a Welsh Language perspective?

From: [WG Official 1]

Sent: 07 September 2026 10:59

To: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Economy Press Office]; [WG Official 5]

Cc: [WG Official 6]; [Education Welsh Language Press Office]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Subject: RE: For input today - PN On Cosine partnership

Diolch [WG Official 2]

[WG Official 3] [WG Official 4] [WG Official 5] – how about something like the below?

The Cabinet Secretary for Economy, Energy and Planning said: “AI has the potential to be of great benefit to the economy and many areas of public life. By working in partnership with companies like Cosine we can make sure those opportunities are realised and that people in Wales can use the technology in their preferred language.”

[WG Official 1]

ATISN 27572 - Email - 20260907 - "For input today - PN On Cosine partnership" –

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 7 Medi 2026 09:46

At: [WG Official 2]; [WG Official 3]; [Economy Press Office]; [WG Official 4]

Copi/Cc: [WG Official 5]; [Education Welsh Lang Press Office]; [WG Official 6]; [WG Official 7]; [WG Official 8]

Pwnc: For input today - PN On Cosine partnership

Bore da

Here's a draft PN on the partnership with Cosine to improve Welsh Language AI. The aim is to send it out this Thursday, for publication on Friday.

[WG Official 5] has already reviewed from Cymraeg policy. Grateful if you could check the draft and work with colleagues in [Economy press office] on lines for CMECE's quote. Can you do this today please so we can send up for ministerial clearance asap?

Diolch yn fawr

[WG Official 1]

ATISN 27572 - Email - 20260814 - "RE: Cosine & Welsh Gvt"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Gwener, 14 Awst 2026 12:35

At: [WG Official 2]; [WG Official 3]; [WG Official 4]; [WG Official 5]

Copi/Cc: [WG Official 6]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Pwnc: RE: Cosine & Welsh Gvt

Thanks [WG Official 3], [WG Official 2], really appreciate you leaning in on this.

Kind regards

[WG Official 1]

From: [WG Official 2]

Sent: 14 August 2026 10:01

To: [WG Official 3]; [WG Official 1]; [WG Official 4]; [WG Official 5]

Cc: [WG Official 6]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Subject: ATB/RE: Cosine & Welsh Gvt

Diolch [WG Official 3].

[WG Official 3], [WG Official 5] er gwyb.

Oddi wrth: [WG Official 3]

Anfonwyd: Dydd Gwener, 14 Awst 2026 09:51

At: [WG Official 1]; [WG Official 2]

Copi/Cc: [WG Official 6]; [WG Official 7]; [WG Official 8]; [WG Official 9]

Pwnc: RE: Cosine & Welsh Gvt

Thanks [WG Official 1].

[WG Official 2] and I caught up briefly yesterday and agreed an outline way forward:

- Cymraeg 2050 to provide a briefing and background to CMEWL on her return from recess, first week of September
- Press release – we'll need to loop in our respective Press teams, who normally lead on these. Likely we'll want to develop a more WG-centric version of the current Cosine draft and co-ordinate a join release/quotes. Work to start asap.

[redacted information]

[WG Official 2] – feel free to cut across any of the above!

[WG Official 3]

ATISN 27572 - Email - 20260805a - "RE: Cosine & Welsh Gvt"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mercher, 5 Awst 2026 16:52

At: [WLGA 1]; [WG Official 2]; [WG Official 7]; [WG Official 3]; [DHCW 1]; [WG Official 4]; [WG Official 5]; [WG Official 6]

Copi/Cc: [WG Official 7]; [Glyn Jones Chief Digital Officer]; [WG Official 8]; [WG Official 9]

Pwnc: RE: Cosine & Welsh Gvt

Hi All,

[redacted information]

Can I also highlight we are currently working with Cosine on a press release relating to our recent collaboration on training their AI models on Welsh language, FYI only at this stage.

Kind regards

[WG Official 1]

From: [WG Official 1]

Sent: 20 July 2026 15:10

To: [WG Official 2]; [WG Official 7]>; [WLGA 1]; [DHCW 1]; [WG Official 3]

Cc: [WG Official 4]; [Glyn Jones Chief Digital Officer]; [WG Official 5]; [WG Official 6]

Subject: FW: Cosine & Welsh Gvt

From: [WG Official 1]

Sent: 20 July 2026 15:10

To: [WG Official 2]; [WG Official 7]>; [WLGA 1]; [DHCW 1]; [WG Official 3]

Cc: [WG Official 4]; [Glyn Jones Chief Digital Officer]; [WG Official 5]; [WG Official 6]

Subject: FW: Cosine & Welsh Gvt

All,

Please forgive the somewhat eclectic distribution list, but I wanted to bring this to your attention as I think it presents an interesting opportunity.

A few weeks ago, I was introduced to Cosine through a mutual contact and, over the last few days, have had discussions with them about potential collaboration, particularly around supporting the training and development of their large language models (LLMs).

Cosine is a UK sovereign AI company backed by the UK Sovereign AI Fund. Interestingly, it was also referenced by the CMECE during our recent discussion on AI in Wales, which was attended by both the CMECE and CMGEC. They are already working with the

Ministry of Defence, collaborating with the Isambard Supercomputing Centre in Bristol, and engaging with NHS England on use cases involving ICD coding.

Today, we initiated working with Cosine to explore the development of enhanced Welsh-language capabilities within their models.

[redacted information]

Please let me know what you think

KR

[WG Official 1]

ATISN 27572 - Email - 20260805 - "RE: Potential Cosine partnership announcement"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mercher, 5 Awst 2026 09:27

At: [Glyn Jones Chief Digital Officer]; [WG Official 2]

Copi/Cc: [WG Official 3]

Pwnc: RE: Potential Cosine partnership announcement

Diolch Glyn, no commercial obligations.

[WG Official 2] shall I drop [Cosine 1] at Cosine a line confirming our interest in a PR?

From: [Glyn Jones Chief Digital Officer]
Sent: 04 August 2026 12:02
To: [WG Official 1]; [WG Official 2]
Cc: [WG Official 3]
Subject: ATB/RE: Potential Cosine partnership announcement

Cytuno hwn yn syniad ardderchog gyda mantais o neges positif i ni hefyd os nad oes unrhyw oblygiadau ariannol/commercial

Cadwch [WG Official 4] yn y loop

Oddi wrth: [WG Official 1]
Anfonwyd: Dydd Iau, 30 Gorffennaf 2026 13:22
At: [WG Official 2]; [Glyn Jones Chief Digital Officer]
Copi/Cc: [WG Official 3]
Pwnc: ATB/RE: Potential Cosine partnership announcement

Diolch [WG Official 2]. Unwaith i Glyn gadarnhau, allen ni ystyried rhywbeth ar y cyd falle?

Oddi wrth: [WG Official 1]
Anfonwyd: Dydd Iau, 30 Gorffennaf 2026 09:24
At: [WG Official 2]; [Glyn Jones Chief Digital Officer]
Copi/Cc: [WG Official 3]
Pwnc: RE: Potential Cosine partnership announcement

Hi [WG Official 2],

Spoke with [WG Official 4] Tuesday and he confirmed that Cosine are very much on the radar of the CMECE [redacted information]. Do you want me to check if his team want to be involved in any PR, subject to green light from Glyn?

Cheers

[WG Official 1]

From: [WG Official 1]
Sent: 30 July 2026 09:18
To: [Glyn Jones Chief Digital Officer]
Cc: [WG Official 2]; [WG Official 3]
Subject: YML/FW: Potential Cosine partnership announcement

Hi Glyn, gw isod. Cyfle i ni gyhoeddi partneriaeth gyda Cosine, bydden ni'n adlewyrchu geiriad partneriaeth OpenAI, rhannu adnoddau ac arbenigedd.

Ti'n hapus i ni fwrw mlaen gyda hwn?

[WG Official 1]

From: [Cosine 1]
Sent: Thursday, July 30, 2026 7:19 am
To: [WG Official 1]; [WG Official 2]
Cc: [Cosine 2]; [Cosine 3]
Subject: Potential Cosine partnership announcement

Hi [WG Official 1], [WG Official 2],

Hope all well! Quick one - which I don't believe we've discussed. Would you be happy for us to work on a joint press release about our engagement with you embedding Welsh language data into our Lumen model training? We'd love to publicise how we're working together.

If this is premature - or you'd rather not for any reason - then absolutely fine.

Best,

[Cosine 1]

ATISN 27572 - Email - 20260731 - "RE: Cosine & Welsh Gvt"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Gwener, 31 Gorffennaf 2026 13:16

At: [WG Official 3]

Copi/Cc: [WG Official 2]

Pwnc: RE: Cosine & Welsh Gvt

Awesome news. FYI we are working on a joint press release with Cosine regarding our new collaboration on Welsh Language models. Also, I spoke to [WG Official 3] earlier this week and he is also supportive and confirmed the CMECE is also keen to meet with them.

I'm speaking to Glyn and our Directorate comms lead Monday, ill come back on next steps.

Thanks both

Have a great weekend.

[WG Official 1]

From: [WG Official 3]
Sent: 31 July 2026 13:01
To: [WG Official 1]
Cc: [WG Official 2]
Subject: RE: Cosine & Welsh Gvt

Hi [WG Official 1]

Sorry to be slow coming back to you – definitely interested. Happy to support – [WG Official 2], copied in, can bring in the wider team.

Thanks

[WG Official 3]

From: [WG Official 1]
Sent: 20 July 2026 15:10
To: [WG Official 2]; [WG Official 7]; [WLGA 1]; [DHCW 1]; [WG Official 3]
Cc: [WG Official 4]; [Glyn Jones Chief Digital Officer]; [WG Official 5]; [WG Official 6]
Subject: FW: Cosine & Welsh Gvt

All,

Please forgive the somewhat eclectic distribution list, but I wanted to bring this to your attention as I think it presents an interesting opportunity.

A few weeks ago, I was introduced to Cosine through a mutual contact and, over the last few days, have had discussions with them about potential collaboration,

particularly around supporting the training and development of their large language models (LLMs).

Cosine is a UK sovereign AI company backed by the UK Sovereign AI Fund. Interestingly, it was also referenced by the CMECE during our recent discussion on AI in Wales, which was attended by both the CMECE and CMGEC. They are already working with the Ministry of Defence, collaborating with the Isambard Supercomputing Centre in Bristol, and engaging with NHS England on use cases involving ICD coding.

Today, we initiated working with Cosine to explore the development of enhanced Welsh-language capabilities within their models.

[redacted information]

Please let me know what you think

KR

[WG Official 1]

[redacted information]

On Wed, 15 Jul 2026 at 12:55, [Cosine 1] wrote:

Hi [WG Official 1],

Great to speak to you today. It was the best meeting of my day (and I promise I don't always say that!)

[redacted information]

- Thank you for the intro to [WG Official 2]. We will take forward

Keep in touch,

[Cosine 1]

ATISN 27572 - Email - 20260715 - "RE: RE: Welsh Language LLM's - Cosine"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mercher, 15 Gorffennaf 2026 15:59

At: [WG Official 2]; [WG Official 3]

Copi/Cc: [Glyn Jones Chief Digital Officer]

Pwnc: RE: RE: Welsh Language LLM's - Cosine

Diolch [WG Official 2], will copy you on my follow up with Cosine.

From: [WG Official 2]

Sent: 15 July 2026 14:48

To: [WG Official 1] ; [WG Official 3]

Cc: [Glyn Jones Chief Digital Officer]

Subject: ATB/RE: RE: Welsh Language LLM's - Cosine

Great, happy to be involved, diolch [WG Official 1].

[WG Official 2]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mercher, 15 Gorffennaf 2026 14:24

At: [WG Official 2]; [WG Official 3]

Copi/Cc: [Glyn Jones Chief Digital Officer]

Pwnc: RE: Welsh Language LLM's - Cosine

Hi [WG Official 2], [WG Official 3],

I had a call with the CRO of Cosine today, (intro via one of my old team). Cosine are a UK based LLM, competitor to Anthropic and Open AI, and is backed by the UK Sov AI Fund. [The Coding Model That Actually Gets It | Cosine](#)

They are keen to work with us directly in training their model in Welsh Language, no fees, but they may want to market the fact they are working with us at some point in the future, which I think has value on both sides, especially if we jointly develop a market leading WL LLM.

We are investigating other collaboration opportunities with Cosine to train their model on specific Welsh Public Sector challenges, [redacted information], but the language angle is something I think we can progress very quickly, hence the email.

Next steps would be a follow up call with their AI team to discuss formal collaboration and how to train their model in Welsh in more detail.

What do you think?

Kind regards

[WG Official 1]

ATISN 27572 – Email – 20260904 – “ATB/RE: Cosine & Welsh Gvt”

From: [WG Official 1]

Sent: 04 September 2026 15:24

To: [WG Official 2]

Subject: ATB/RE: Cosine & Welsh Gvt

Diolch [WG Official 2], wedi tracio un awgrym. Ti'n hapus i rannu hwn gyda [WG Official 3] Swyddfa AI (ar wylie ahob ond nol d llun) cyn i ti ddanfon at swyddfa CMECE plîs?

Dwi'n briffio CMEWL ar 9fed felly *ideally bydde hwn yn mynd lan ar ôl hwnna fel bod dim crossed wires ond dwed os nad yw hwnna'n ymarferol.

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Gwener, 4 Medi 2026 15:13

At: [WG Official 2]

Pwnc: RE: Cosine & Welsh Gvt

Haia

Nag oes. Ond dwi wedi paratoi drafft o'r datganiad i'r wasg. Fedri di fwrw golwg a chywiro / newid cyn i mi ei anfon at swyddfa'r wasg [WG Official 3] am ddyfyniad CMECE plis?

Mae'r stori wedi ei amserlennu i fynd allan ddydd lau nesa dan embargo tan fore Gwener (11 Medi); felly bydd angen i ni gytuno'r drafft efo Cosine cyn hynny hefyd.

Hwyl!

[WG Official 1]

From: [WG Official 1]

Sent: 04 September 2026 15:02

To: [WG Official 2]

Subject: Re: Cosine & Welsh Gvt

Hi [WG Official 2],

Oes unrhywun o'r swyddfa AI wedi cysylltu da ti am baratodau PR? (Ma lot wedi bodi ffwrdd so dim panics os ddim) Ma cyfarfod gyda Cosine diwedd bore Llun so just ishe checko.

Diolch eto am yr help gyda hwn.

[WG Official 1]

From: [WG Official 1]
Sent: Tuesday, August 25, 2026 2:37 pm
To: [WG Official 2]; [WG Official 3]
Cc: [WG Official 4]; [WG Official 5]
Subject: ATB/RE: Cosine & Welsh Gvt

Adding [WG Official 3] - some background info for the Cosine Partnership press release.

Technoleg ac AI Cymraeg - briff CMEWL.docx

Welsh Technology and AI brief (EN).docx

[WG Official 1]

Oddi wrth: [WG Official 1]
Anfonwyd: Dydd Mawrth, 25 Awst 2026 10:34
At: [WG Official 2]
Copi/Cc: [WG Official 3]; [WG Official 4]
Pwnc: ATB/RE: Cosine & Welsh Gvt

Thanks [WG Official 2].

Added a couple of sentences on sovereign AI on [WG Official 5]'s request – do you have any lines on sovereign AI I could add to her general briefing going forward?

Thanks,

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mawrth, 25 Awst 2026 08:38

At: [WG Official 2]

Copi/Cc: [WG Official 3]; [WG Official 4]

Pwnc: RE: Cosine & Welsh Gvt

Hi [WG Official 2]

One comment and some minor updates on the [redacted information] – all tracked.

Feel free to tweak as needed.

[WG Official 1]

From: [WG Official 1]

Sent: 21 August 2026 15:22

To: [WG Official 2]

Cc: [WG Official 3]; [WG Official 4]

Subject: YML/FW: Cosine & Welsh Gvt

Hi [WG Official 2],

It's looking likely that we'll be briefing CMEWL on this (and wider Welsh tech developments) on the 9th of September.

Drafted a quick brief which will be the basis of that DC - are you happy with the wording, **[redacted information]**?

Welsh Technology and AI brief (EN).docx

Diolch,

[WG Official 1]

ATISN 27572 - Email - 20260911 - "Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement"

Oddi wrth: [Cosine 1]

Anfonwyd: Dydd Gwener, 11 Medi 2026 12:07

At: [WG Official 1]

Copi/Cc: [Cosine 2]; [Cosine 3]; [Cosine 4]; [WG Official 2]; [WG Official 3]; [Education Welsh Language Press Office]; [WG Official 4]

Pwnc: Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Excellent!

Here is a link to our post as well. We look forward to working with you!

Best,

[Cosine 1]

On Fri, 11 Sept 2026 at 09:32, [WG Official 1] wrote:

Bore da,

This press release has now been issued to media with no embargo.

Diolch,

[WG Official 1]

ATISN 27572 - Email - 20260910b - "RE: ATB/RE: ATB/RE: Potential Cosine partnership announcement"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Iau, 10 Medi 2026 17:01

At: [Cosine 1]

Copi/Cc: [WG Official 2]; [Cosine 2]; [WG Official 3]; [Cosine 3]; [Cosine 4]

Pwnc: RE: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Apologies – this is the right version (attached).

[WG Official 1]

From: [WG Official 1]

Sent: 10 September 2026 16:28

To: [Cosine 1]

Cc: [WG Official 2]; [Cosine 2]; [WG Official 3]; [Cosine 3]; [Cosine 4]; [Education Welsh Language Press Office]

Subject: RE: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Thank you very much.

We've now had it cleared our end (apologies for the delay). As it's quite late in the day we'll issue it straight tomorrow morning rather than under embargo tonight. It will be at

9:30am. I'll be out of the office tomorrow but I've copied in my press office colleagues who can let you know as soon as it's out. Attached for info.

Best wishes

[WG Official 1]

ATISN 27572 - Email - 20260910a - "Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement"

Oddi wrth: [Cosine 1]

Anfonwyd: Dydd Iau, 10 Medi 2026 16:23

At: [WG Official 1]

Copi/Cc: [WG Official 2]; [Cosine 2]; [WG Official 3]; [Cosine 3]; [Cosine 4]

Pwnc: Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 1],

I've attached a few versions of the logo that you can use!

Thanks,

[Cosine 1]

On Thu, 10 Sept 2026 at 11:45, [WG Official 1] wrote:

Hi [Cosine 1]

We're just waiting for clearance this end, and will let you know asap when that's done.

In the meantime, I'd like to get things ready on the system. Do you have any pictures or even a logo (landscape format) we could use to go with the press release please?

Thanks

[WG Official 1]

From: [WG Official 1]

Sent: 10 September 2026 10:59

To: [Cosine 1]; [WG Official 2]

Cc: [Cosine 2]; [WG Official 3]; [Cosine 3]; [Cosine 4]

Subject: ATB/RE: ATB/RE: ATB/RE: Potential Cosine partnership announcement

[WG Official 2] could you confirm please?

Diolch,

[WG Official 1]

Oddi wrth: [Cosine 1]

Anfonwyd: Dydd Iau, 10 Medi 2026 10:46

At: [WG Official 1]

Copi/Cc: [Cosine 2]; [WG Official 2]; [Cosine 3]; [Cosine 4]

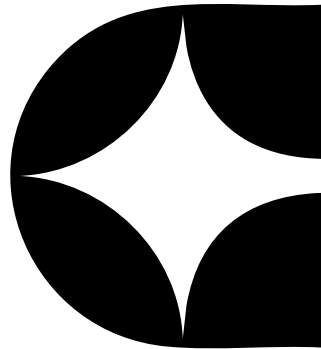
Pwnc: Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 1],

Can you confirm what time the embargo lifts and you are planning your announcement?
We are trying to coordinate ways to promote this more broadly and will ideally promote this in the early afternoon.

Thanks!

[Cosine 1]



ATISN 27572 - Email - 20260907 - "Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement"

Oddi wrth: [Cosine 1] <[Cosine 1]>

Anfonwyd: Dydd Llun, 7 Medi 2026 15:09

At: [WG Official 1]

Copi/Cc: [Cosine 2]; [WG Official 2]; [Cosine 3]; [Cosine 4]

Pwnc: Re: ATB/RE: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 1],

That sounds great! I've confirmed that this will work in our schedule as well so we'll plan for an announcement on 11 September.

[redacted information]

Thanks,

[Cosine 1]

On Mon, 7 Sept 2026 at 13:19, [WG Official 1] wrote:

Great, thanks [Cosine 1].

Our Press Office have drafted a Welsh Government press release and included the CEO quote from your release, with the intention of having identical quotes appear in both announcements.

Thanks again,

[WG Official 1]

Oddi wrth: [Cosine 1] <[Cosine 1]>

Anfonwyd: Dydd Llun, 7 Medi 2026 13:01

At: [WG Official 1]

Copi/Cc: [Cosine 2]; [WG Official 2]; [Cosine 3]; [Cosine 4]

Pwnc: Re: ATB/RE: Potential Cosine partnership announcement

Thank you [WG Official 1]! Looping in [Cosine 4] from our content team regarding the schedule and confirm shortly.

I'm assuming you're working off the existing draft but please let me know if you need any updated copy from our side!

Thanks!

[Cosine 1]

On Mon, 7 Sept 2026 at 12:23, [WG Official 1] wrote:

Hi [Cosine 1],

Thanks for your patience over the last few weeks. We're working towards a press release for Friday 11 September so will share Ministerial quotes with you Thursday at the very latest. We're expecting two quotes, one from the Cabinet Minister for Education and the Welsh Language and one from the Cabinet Minister for Enterprise, Connectivity and Energy.

Hope that timing works for you but any issues in the meantime let me know.

Thanks,

[WG Official 1]

ATISN 27572 - Email - 20260813 - "Re: Potential Cosine partnership announcement"

Oddi wrth: [Cosine 1] <[Cosine 1]>

Anfonwyd: Dydd Iau, 13 Awst 2026 09:45

At: [WG Official 1]

Copi/Cc: [WG Official 2]; [WG Official 3]; [Cosine 2]; [WG Official 4]; [Cosine 3]

Pwnc: Re: Potential Cosine partnership announcement

Great, thank you for the update! Please let me know if it would be helpful to hop on a call in the meantime.

Thanks,

[Cosine 1]

On Thu, 13 Aug 2026 at 16:38, <[WG Official 1]> wrote:

Thanks [Cosine 1]. Mid-sept is probably the earliest we could do but we can work on the draft asap so that everything's ready to go.

[WG Official 2] er gwyb.

[WG Official 1]

From: [Cosine 1] <[Cosine 1]>

Sent: Thursday, August 13, 2026 4:51 am

To: [WG Official 1]

Cc: [Cosine 2] <[Cosine 2]>; [WG Official 4]; [Cosine 3] <[Cosine 3]>

Subject: Re: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 1],

We'd be happy to do an earlier announcement! I know we are highlighting the model's Welsh language capabilities in our first newsletter this week so it would be great to pair an announcement with that.

If we need more time for approvals, I'm happy to schedule it for Mid-September though. I've taken a first pass at a PR draft but please let me know if there is context in the partnership I'm missing.

Here is a google doc version but shared in .docx as well.

Thanks,

[Cosine 1]

On Wed, 12 Aug 2026 at 23:44, [WG Official 1] wrote:

Thanks [Cosine 1], sounds good. Could we work towards a mid-Sept announcement?

And yes please, grateful if you could share a draft press release with us.

Thanks,

[WG Official 1]

Oddi wrth: [Cosine 1] <[Cosine 1]>

Anfonwyd: Dydd Mercher, 12 Awst 2026 15:38

At: [Cosine 2] <[Cosine 2]>

Copi/Cc: [WG Official 4]; [WG Official 1]; [Cosine 3]

Pwnc: Re: Potential Cosine partnership announcement

Hi [WG Official 4],

[redacted information] We would like to start the relationship with an announcement of the relationship. Are you able to help us get that pulled together?

Happy to share a draft press release for you all to edit, if that would be helpful!

Best,

[Cosine 1]

On Sat, 8 Aug 2026 at 00:26, [Cosine 2] <[Cosine 2]> wrote:

Hi [WG Official 4],

Great news! As it happens the very very first iteration of our model went live about an hour ago, and the second request was to ask it to say happy birthday in Welsh... so I can confirm that it is working!

Are you free from 10:30am-12pm on Tuesday for a call?

[Cosine 2]

On Fri, 7 Aug 2026 at 16:31, <[WG Official 4]> wrote:

Hi [Cosine 2],

Hope your well. Great news, we have been given the green light to progress discussions regarding a joint press release. Shall we have a brief chat regarding next steps at some point next week?

Have a great weekend

[WG Official 4]

From: [Cosine 2] <[Cosine 2]>

Sent: 30 July 2026 07:19

To: [WG Official 4]; [WG Official 1]

Cc: [Cosine 1] <[Cosine 1]>; [Cosine 3]

Subject: Potential Cosine partnership announcement

Hi [WG Official 4], [WG Official 1],

Hope all well! Quick one - which I don't believe we've discussed. Would you be happy for us to work on a joint press release about our engagement with you embedding Welsh language data into our Lumen model training? We'd love to publicise how we're working together.

If this is premature - or you'd rather not for any reason - then absolutely fine.

Best,

[Cosine 2]

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the

ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

ATISN 27572 - Email - 20260813 - "FW: ATB: RE: Potential Cosine partnership announcement"

Oddi wrth: [WG Official 1] <[WG Official 1]>

Anfonwyd: Dydd Iau, 13 Awst 2026 14:57

At: [WG Official 2]; [WG Official 3]; [WG Official 4]

Pwnc: FW: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 2],

As discussed, please find below and attached the info on the PR with Cosine relating to our collaboration on training their emerging LLM in Welsh. We have had the green light

from Glyn, but we may want sign-off and inputs from the CMECE and CMEWL, so will need to go through the proper process.

I assume it's easier we send this up to both private offices as one briefing, with inputs and support from [WG Official 4] and team, but will be guided by your experience on these matters.

[WG Official 3] are you happy for our team to take the lead on the internal approval processes with your support and then work jointly on the content of the PR itself, I assume with support from ECE comms.

[redacted information]

Many thanks

[WG Official 1]

From: [Cosine 1]

Sent: 13 August 2026 04:51

To: [WG Official 3]

Cc: [Cosine 2]; [WG Official 1]; [Cosine 3]

Subject: Re: ATB/RE: Potential Cosine partnership announcement

Hi [WG Official 3],

We'd be happy to do an earlier announcement! I know we are highlighting the model's Welsh language capabilities in our first newsletter this week so it would be great to pair an announcement with that.

If we need more time for approvals, I'm happy to schedule it for Mid-September though. I've taken a first pass at a PR draft but please let me know if there is context in the partnership I'm missing.

Here is a google doc version but shared in .docx as well.

Thanks,

[Cosine 1]

On Wed, 12 Aug 2026 at 23:44, [WG Official 3] wrote:

Thanks [Cosine 1], sounds good. Could we work towards a mid-Sept announcement?

And yes please, grateful if you could share a draft press release with us.

Thanks,

[WG Official 3]

Oddi wrth: [Cosine 1] <[Cosine 1]>

Anfonwyd: Dydd Mercher, 12 Awst 2026 15:38

At: [Cosine 2] <[Cosine 2]>

Copi/Cc: [WG Official 1] <[WG Official 1]>; [WG Official 3]; [Cosine 3]

Pwnc: Re: Potential Cosine partnership announcement

Hi [WG Official 1],

[redacted information] We would like to start the relationship with an announcement of the relationship. Are you able to help us get that pulled together?

Happy to share a draft press release for you all to edit, if that would be helpful!

Best,

[Cosine 1]

On Sat, 8 Aug 2026 at 00:26, [Cosine 2] <[Cosine 2]> wrote:

Hi [WG Official 1],

Great news! As it happens the very very first iteration of our model went live about an hour ago, and the second request was to ask it to say happy birthday in Welsh... so I can confirm that it is working!

Are you free from 10:30am-12pm on Tuesday for a call?

[Cosine 2]

On Fri, 7 Aug 2026 at 16:31, <[WG Official 1]> wrote:

Hi [Cosine 2],

Hope your well. Great news, we have been given the green light to progress discussions regarding a joint press release. Shall we have a brief chat regarding next steps at some point next week?

Have a great weekend

[WG Official 1]

From: [Cosine 2] <[Cosine 2]>

Sent: 30 July 2026 07:19

To: [WG Official 1] <[WG Official 1]>; [WG Official 3]

Cc: [Cosine 1] <[Cosine 1]>; [Cosine 3]

Subject: Potential Cosine partnership announcement

Hi [WG Official 1], [WG Official 3],

Hope all well! Quick one - which I don't believe we've discussed. Would you be happy for us to work on a joint press release about our engagement with you embedding Welsh language data into our Lumen model training? We'd love to publicise how we're working together.

If this is premature - or you'd rather not for any reason - then absolutely fine.

Best,

[Cosine 2]

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

Sganiwyd y neges hon am bob feirws hysbys wrth iddi adael Llywodraeth Cymru. Mae Llywodraeth Cymru yn cymryd o ddifrif yr angen i ddiogelu eich data. Os cysylltwch â Llywodraeth Cymru, mae ein hysbysiad preifatrwydd yn esbonio sut rydym yn defnyddio eich gwybodaeth a sut rydym yn diogelu eich preifatrwydd. Rydym yn croesawu

gohebiaeth yn Gymraeg. Byddwn yn anfon ateb yn Gymraeg i ohebiaeth a dderbynnir yn Gymraeg ac ni fydd gohebu yn Gymraeg yn arwain at oedi. ----- On leaving the Welsh Government this email was scanned for all known viruses. The Welsh Government takes the protection of your data seriously. If you contact the Welsh Government then our Privacy Notice explains how we use your information and the ways in which we protect your privacy. We welcome receiving correspondence in Welsh. Any correspondence received in Welsh will be answered in Welsh and corresponding in Welsh will not lead to a delay in responding.

ATISN 27572 - Email - 20260807 - "RE: Potential Cosine partnership announcement"

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Gwener, 7 Awst 2026 16:31

At: [Cosine 1]; [WG Official 1]

Copi/Cc: [Cosine 2]; [Cosine 3]

Pwnc: RE: Potential Cosine partnership announcement

Hi [Cosine 1],

Hope your well. Great news, we have been given the green light to progress discussions regarding a joint press release. Shall we have a brief chat regarding next steps at some point next week?

Have a great weekend

[WG Official 1]

From: [Cosine 1]
Sent: 30 July 2026 07:19
To: [WG Official 1]
Cc: [Cosine 2]; [Cosine 3]
Subject: Potential Cosine partnership announcement

Hi [WG Official 1],

Hope all well! Quick one - which I don't believe we've discussed. Would you be happy for us to work on a joint press release about our engagement with you embedding Welsh language data into our Lumen model training? We'd love to publicise how we're working together.

If this is premature - or you'd rather not for any reason - then absolutely fine.

Best,

[Cosine 1]

ATISN 27572 - Email - 20260923 - "CS1551 At Eich Sylw / For Your Attention"

Oddi wrth: [WG Official 1]
Anfonwyd: Dydd Mercher, 23 Medi 2026 16:37
At: [WG Official 2]; [WG Official 3]; [WG Official 4]
Copi/Cc: [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]; [WG Official 4]
Pwnc: ATB/RE: CS1551 At Eich Sylw / For Your Attention

Hi [WG Official 2],

Dyma ymateb i CS1551 wedi atodi.

Diolch,

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 21 Medi 2026 09:15

At: [WG Official 2]; [WG Official 3]; [WG Official 4]; [WG Official 5]

Copi/Cc: [WG Official 1]

Pwnc: ATB/RE: CS1551 At Eich Sylw / For Your Attention

Diolch [WG Official 2] – wnai ddiweddarau ein log cwynion i adlewyrchu hynny.

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 21 Medi 2026 09:12

At: [WG Official 2]; [WG Official 3]; [WG Official 4]

Copi/Cc: [WG Official 5]

Pwnc: ATB/RE: CS1551 At Eich Sylw / For Your Attention

Hi

[WG Official 3] yn delio

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Gwener, 18 Medi 2026 10:48

At: [WG Official 2]; [WG Official 3]; [WG Official 4]

Copi/Cc: [WG Official 5]

Pwnc: YML/FW: CS1551 At Eich Sylw / For Your Attention

Helo dîm,

Mae cwyn wedi dod via swyddfa'r Comisiynydd am y cyhoeddiad Cosine i chi weld – os fyddai sgwrs yn ddefnyddiol o ran ymateb, rhowch wybod. Angen ymateb cyn 02/10.

[WG Official 1]

Oddi wrth: [CYG-WLC 1]

Anfonwyd: Dydd Gwener, 18 Medi 2026 10:36

At: [WG Official 1]; [WG Official 2]; [WG Official 3]

Pwnc: CS1551 At Eich Sylw / For Your Attention

Annwyl Swyddog,

CS1551: Cwyn safonau'r Gymraeg

Rydym wedi derbyn cwyn gan berson ar 17/09/2026. Mae'n ymwneud â chyhoeddiad Llywodraeth Cymru am ei chydweithrediad â Cosine. Mae'r achwynydd yn pryderu y bydd hyn yn "tanseilio budd y cyhoedd a chydaddoldeb ieithyddol hirdymor". Isod, dyfynnaf dri phwynt y mae'r achwynydd yn seilio ei gŵyn arnynt:

- Safonau Llundio Polisi (Safonau 88–93):** Methiant Llywodraeth Cymru i gynnal asesiad effaith iaith digonol cyn cytuno i drosglwyddo data cyhoeddus a chorfforaethau iaith i gwmni masnachol preifat (Cosine). Mae perygl difrifol y bydd y penderfyniad hwn yn cael effaith andwyol ar gyfleoedd pobl i ddefnyddio'r Gymraeg, drwy greu monopolies masnachol a thalu ffioedd tanysgrifio/API am adnoddau a gafodd eu hariannu'n gyhoeddus.
- Egwyddor Trin y Gymraeg Ddim Llai Ffafirol na'r Saesneg:** Drwy beidio â sicrhau bod pwysau'r model (model weights) a'r cod yn agored (open-source / open-weights), mae Llywodraeth Cymru yn peryglu mynediad agored sefydliadau cyhoeddus, ysgolion, a llywodraeth leol at dechnoleg iaith Gymraeg sofran, tra bod adnoddau Saesneg enfawr eisoes ar gael am ddim yn fyd-eang.

3. ***Mynediad at Wasanaethau Cyhoeddus a Chyllid Trethdalwyr:*** *Mae'r cytundeb hwn yn amgáu comin diwylliannol cyhoeddus mewn cynnyrch caeedig, gan fygwth gallu cyrff cyhoeddus yng Nghymru i gydymffurfio â'u Safonau eu hunain yn y dyfodol heb orfod talu cwmni masnachol yn Llundain.*

Mae'r gŵyn yn ddilys sy'n golygu bod dyletswydd arnom i ystyried cynnal ymchwiliad dan adran 71 Mesur y Gymraeg.

Er mwyn gwneud yn siŵr bod gennym yr holl wybodaeth sydd ei angen, a dealltwriaeth lawn o'r sefyllfa cyn penderfynu, mae'n bwysig eich bod yn rhoi gymaint o wybodaeth a phosib i ni ar y cam hwn. Gall rhoi digon o wybodaeth i ni nawr olygu na fydd angen i ni gysylltu â chi eto am fwy o wybodaeth pe baem yn penderfynu ymchwilio.

A wnewch chi felly ymateb i'r e-bost hwn yn cadarnhau eich bod yn derbyn cyfrifoldeb am y mater dan sylw, gan roi gwybod i ni am:

- unrhyw amgylchiadau perthnasol sy'n esbonio beth sydd wedi digwydd
- unrhyw gamau rydych wedi eu cymryd yn barod i gywiro'r mater
- unrhyw gamau y byddwch yn eu cymryd yn y dyfodol i geisio gwneud yn siŵr na fydd sefyllfa tebyg yn codi eto.

Anfonwch eich ymateb erbyn 02/10/2026 os gwelwch yn dda. Gallwch ofyn am fwy o amser os oes angen.

Dear Officer,

CS1551 - Welsh language standards complaint

We have received a complaint from a person on 17/09/2026. The complaint concerns an announcement by the Welsh Government regarding its collaboration with Cosine. The complainant is concerned that this will “undermine the public interest and long-term linguistic equality”. Below, I quote three points on which the complainant bases their complaint:

1. ***Policy Making Standards (Standards 88–93):*** *The Welsh Government failed to carry out an adequate Welsh language impact assessment before agreeing to transfer public data and language assets to a private commercial company*

(Cosine). There is a serious risk that this decision will have an adverse effect on people's opportunities to use Welsh, by creating commercial monopolies and imposing subscription/API fees for resources that were publicly funded.

2. ***The Principle of Treating Welsh No Less Favourably than English:*** *By failing to ensure that the model weights and code are open (open-source/open-weights), the Welsh Government is jeopardising open access for public bodies, schools, and local government to sovereign Welsh-language technology, while vast English-language resources are already freely available worldwide.*
3. ***Access to Public Services and Use of Taxpayers' Money:*** *This agreement encloses a public cultural commons within a closed commercial product, threatening the ability of public bodies in Wales to comply with their own Standards in the future without having to pay a commercial company based in London.*

The complaint is valid, which means we are required to consider conducting an investigation under section 71 of the Welsh Language Measure.

To ensure we have all the necessary information and a full understanding of the situation before making a decision, it's important that you provide us with as much information as possible at this stage. Giving us enough information now may mean we won't need to contact you again for more details if we decide to investigate.

Please respond to this e-mail confirming that you take responsibility for the matter in question and let us know about:

- any relevant circumstances that explain what has happened
- any steps you have already taken to correct the matter
- any steps you will take in the future to try to ensure a similar situation does not arise again

Please send your response by 02/10/2026. You can request more time if needed.

Yn gywir,

[CYG-WLC 1]

Dylai'r datganiadau/sylwadau uchod gael eu hystyried yn rhai personol ac nid o reidrwydd yn ddatganiadau neu sylwadau gan Gomisiynydd y Gymraeg, unrhyw ran ohono neu unrhyw gorff sy'n gysylltiedig ag ef. Sganiwyd y neges yma am bob feirws hysbys wrth iddo adael Comisiynydd y Gymraeg. Defnyddir systemau diogelwch Microsoft Enterprise i gyflawni'r gwasanaeth hwn.

The statements/comments made above should be regarded as personal and not necessarily those of the Welsh Language Commissioner, any constituent part or connected body. On leaving the Welsh Language Commissioner this email was scanned for all known viruses. Microsoft Enterprise security applications are used to provide this service.



Ysgrifennwch ataf yn Gymraeg neu Saesneg
Please write to me in Welsh or English

ATISN 27572 – Email - 20260902 – “ATB/RE: Tech briefing”

From: [WG Official 1]

Sent: 02 September 2026 13:47

To: [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a’r Gymraeg]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a’r Gymraeg]; [WG Official 2]

Cc: [Cymraeg 2050 Team]

Subject: ATB/RE: Tech briefing

Hi [WG Official 3],

Please find attached a briefing on the background and upcoming announcements.

Diolch,

[WG Official 1]

Oddi wrth: [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a’r Gymraeg]

Anfonwyd: Dydd Gwener, 21 Awst 2026 15:34

At: [WG Official 1]; [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a’r Gymraeg]; [WG Official 2]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a’r Gymraeg]

Copi/Cc: [Cymraeg 2050 Team]

Pwnc: RE: Tech briefing

Prynhawn da,

I have sent through the diary marker.

Please could we be in receipt of all relevant papers by 15:00pm on the 2.9.26 if possible.

Cofion cynnes

[WG Official 3]

From: [WG Official 1]

Sent: 21 August 2026 14:50

To: [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [WG Official 2]

Cc: [Cymraeg 2050 Team]

Subject: ATB/RE: Tech briefing

Hi [WG Official 3],

Yes, 12:30-13:00 on 9.9.26 suits, diolch.

Could you also include [WG Official 2] please?

Diolch,

[WG Official 1]

Oddi wrth: [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]

Anfonwyd: Dydd Gwener, 21 Awst 2026 14:43

At: [WG Official 1]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]

Copi/Cc: [Cymraeg 2050 Team]

Pwnc: RE: Tech briefing

Prynhawn da,

There is availability in the Cabinet Ministers diary on the 9.9.26 from 12:30-13:00 if the time and date suits.

Cofion cynnes

[WG Official 2]

From: [WG Official 1]

Sent: 21 August 2026 11:02

To: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]

Cc: [Cymraeg 2050 Team]

Subject: ATB/RE: Tech briefing

Diolch [WG Official 2].

We're aiming for the first announcement w/c 14 September.

Thanks again,

[WG Official 1]

Oddi wrth: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]

Anfonwyd: Dydd Gwener, 21 Awst 2026 09:58

At: [WG Official 1]; [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]

Copi/Cc: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [Cymraeg 2050 Team]

Pwnc: ATB/RE: Tech briefing

Bore da pawb,

Just coming in here to highlight that if possible the meeting should be before the announcements mid-September. [WG Official 1]'s reassured me that they would only need 30 minutes with the Minister, so grateful if this can be fit into her busy September diary!

[WG Official 1] could you confirm which exact date the First announcement will be to help diary find a slot beforehand ogydd?

Diolch,

[WG Official 2]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Iau, 20 Awst 2026 14:27

At: [DS Cabinet Minister for Education and the Welsh Language/YD Gweinidog Cabinet dros Addysg a'r Gymraeg]

Copi/Cc: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [Cymraeg 2050 Team]

Pwnc: Tech briefing

Prynhawn da,

As discussed with [WG Official 2] this morning, there are a series of Welsh technology and AI related announcements coming up in the next few months, the first of which is expected mid-September.

Appreciating how busy diaries must be, would it be possible to arrange some time with the Minister before then so that we can brief her on these developments and the wider Welsh tech background?

Diolch,

[WG Official 1]

ATISN 27572 – Email – 20260908 – “ATB/RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept”

From: [WG Official 1] <[WG Official 1]>

Sent: 08 September 2026 16:47

To: [WG Official 2] <[WG Official 2]>; [WG Official 3] <[WG Official 3]>

Subject: ATB/RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Dyma chi, wedi cytuno gyda Swyddfa AI

- Rydym wedi ymrwymo i sicrhau bod AI yn cael ei fabwysiadu mewn ffyrdd diogel, moesegol a thryloyw, gyda chefnogaeth llywodraethu briodol, mesurau diogelu a goruchwyliaeth ddynol.
- Rydym yn datblygu fframwaith llywodraethu gweithredol cenedlaethol ar gyfer AI i helpu gwasanaethau cyhoeddus Cymru i fabwysiadu AI yn gyfrifol a rhoi'r eglurder a'r sicrwydd sydd eu hangen arnynt.
- Mae Cynllun AI i Gymru yn parhau i fod yn fframwaith strategol ar gyfer mabwysiadu AI yn gyfrifol. Mae ein ffocws bellach ar weithredu, gan ddarparu'r gefnogaeth ymarferol angenrheidiol i helpu sefydliadau i fabwysiadu AI yn ddiogel, yn foesegol ac yn effeithiol.

Oddi wrth: [WG Official 1] <[WG Official 1]>

Anfonwyd: Dydd Llun, 7 Medi 2026 17:45

At: [WG Official 2] <[WG Official 2]>

Copi/Cc: [WG Official 3] <[WG Official 3]>

Pwnc: ATB/RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Gret - diolch. Oes leins da ni am ethics hyn oll os yw rhai yn codi?

[WG Official 1]

Oddi wrth: [WG Official 1] <[WG Official 1]>

Anfonwyd: Dydd Llun, 7 Medi 2026 16:49

At: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Copi/Cc: [WG Official 2] <[WG Official 2]>; [Education Welsh Language Press Office]; [WG Official 3] <[WG Official 3]>; [WG Official 4] <[WG Official 4]>; [WG Official 5] <[WG Official 5]>; [WG Official 6] <[WG Official 6]>; [WG Official 7] <[WG Official 7]>; [WG Official 8] <[WG Official 8]>; [Economy Press Office]; [WG Official 9] <[WG Official 9]>; [WG Official 10] <[WG Official 10]>; [WG Official 11] <[WG Official 11]>; [WG Official 12] <[WG Official 12]>; [WG Official 13] <[WG Official 13]>; [WG Official 14] <[WG Official 14]>; [WG Official 15] <[WG Official 15]>; [WG Official 16] <[WG Official 16]>; [WG Official 17] <[WG Official 17]>; [WG Official 18] <[WG Official 18]>; [WG Official 19] <[WG Official 19]>; [WG Official 20] <[WG Official 20]>; [WG Official 21] <[WG Official 21]>; [WG Official 22] <[WG Official 22]>; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]

Pwnc: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Hello

Please find attached a draft PN on a new partnership with AI company Cosine to advance Welsh language AI.

Can we get clearance by both CMs by 5pm on Wednesday please?

Diolch yn fawr

[WG Official 1]

ATISN 27572 – Email – 20260914 – “Lines on Cosine”

From: [WG Official 1]

Sent: 14 September 2026 16:30

To: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a’r Gymraeg]; [LCC - Government Business / Busnes y Llywodraeth]

Cc: [WG Official 2]; [WG Official 3]

Subject: Lines on Cosine

Helo,

Diolch yn fawr am y sgwrs prynhawn ma.

As discussed, lines on the new partnership with Cosine, and ethical use of AI, to be added to tomorrow’s brief if possible.

Diolch,

[WG Official 1]

Partneriaeth Cosine:

- Mae'r bartneriaeth newydd gyda Cosine yn helpu i gryfhau gallu Cymraeg mewn AI, gan greu mwy o gyfleoedd i ni i gyd ddefnyddio ein Cymraeg yn hyderus bob dydd a chefnogi dyfodol yr iaith yn yr oes ddigidol.
- Nid oes ymrwymiad ariannol i Lywodraeth Cymru; rhannu arbenigedd ac adnoddau **agored** yn unig fydd sail y bartneriaeth hon.

- Mae'r Gymraeg yn perthyn i ni i gyd a ni ddylai fod unrhyw rwystr ariannol i ddefnyddio'r Gymraeg. Mae'r holl adnoddau technoleg Cymraeg sy'n cael ei ariannu gan Lywodraeth Cymru ar gael yn rhad ac am ddim er mwyn i **unrhyw un** eu defnyddio a'u hailddefnyddio.

AI moesegol:

- Rwy'n gwybod bod rhybuddion wedi bod yn y cyfryngau yn ddiweddar am AI. Rydym yn monitro ac yn adolygu'r sefyllfa AI y DU a byd-eang yn barhaus.
- Rydym wedi ymrwymo i sicrhau bod AI yn cael ei fabwysiadu mewn ffyrdd diogel, moesegol a thryloyw, gyda goruchwyliaeth ddynol briodol a chyfrifoldeb clir am wneud penderfyniadau.
- Rydym yn datblygu fframwaith llywodraethu gweithredol cenedlaethol ar gyfer AI i helpu gwasanaethau cyhoeddus Cymru i fabwysiadu AI yn gyfrifol.

Dolenni perthnasol:

Llywodraeth Cymru a Cosine yn cydweithio i gryfhau deallusrwydd artiffisial Cymraeg | LLYW.CYMRU

AI staff 'genuinely frightened' for humanity's future, ex-Anthropic researcher tells BBC - BBC News

MPs and Lords call for new law to address AI threat to human rights - BBC News

Anthropic boss Dario Amodei calls for AI development to slow down - BBC News

ATISN 27572 – Email – 20260922 - YML/FW: Leins Ychwanegol Cosine

From: [WG Official 1]

Sent: 22 September 2026 12:02

To: [WG Official 2]

Subject: YML/FW: Leins Ychwanegol Cosine

Er gwyb

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mawrth, 22 Medi 2026 11:33

At: [WG Official 2]

Copi/Cc: [WG Official 3]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [WG Official 4]

Pwnc: YML/FW: Leins Ychwanegol Cosine

Helo

Cais gan PS i ddiweddarau'r A-Y gyda'r rhain yn hot topics plîs. Ti folon neud?

Diolch!

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mawrth, 22 Medi 2026 11:29

At: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]

Copi/Cc: [WG Official 2]; [Cymraeg 2050]; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]; [WG Official 4]; [Glyn Jones Chief Digital Officer]; [WG Official 6]; [WG Official 7]; [WG Official 8]

Pwnc: Leins Ychwanegol Cosine

Helo'na,

Dyma wybodaeth ychwanegol ar Cosine yn unol â'ch cais.

Cofion gorau

[WG Official 1]

ATISN 27572 – Email – 20260924 – “BIRT”

From: [WG Official 1]

Sent: 24 September 2026 14:39

To: [WG Official 2]

Subject: BIRT

Hapus gyda hwn?

[ID] Gweithio gyda’r sector technoleg er mwyn sicrhau’r defnydd o’r cydrannau technoleg iaith // Work with the technology sector to ensure use of the language technology Components

Mae’r gwaith isod yn parhau. Yn fwyaf diweddar, cyhoeddwyd partneriaeth newydd gyda Cosine, cwmni AI sofran, i helpu i wella sut mae technoleg AI yn prosesu’r Gymraeg. Yn ehangach na hyn, yn dilyn llythyrau gan y Gweinidog at gwmnïau technoleg rhyngwladol, rydym bellach yn trefnu’r cyfarfodydd hynny i drafod sut y gellir cryfhau’r gefnogaeth i’r Gymraeg yn eu hadnoddau. O dan ein partneriaeth bresennol gyda Microsoft, bydd ffocws penodol y cyfarfod hwnnw’n ymwneud â **[redacted information]**.

Mae’r grantiau technoleg ar gyfer 2026-27 (£406,400) bellach yn mynd rhagddynt hefyd, gan gefnogi’r gwaith o gynnal a datblygu gwasanaethau ac adnoddau technoleg Cymraeg. Mae astudiaethau defnydd newydd bellach wedi’u cyhoeddi ar dudalen *Technoleg Cymraeg a deallusrwydd artiffisial (AI): diweddariadau | LLYW.CYMRU*, gan ddangos sut mae ein dull o ddatblygu a rhannu adnoddau technoleg Cymraeg yn agored ac am ddim yn gweithio yn ymarferol. Wrth i fwy o bobl a sefydliadau ddefnyddio’r adnoddau agored hyn, byddwn yn parhau i ychwanegu rhagor o astudiaethau defnydd. Rydym yn parhau i ddiweddarau ein rhestr adnoddau technoleg Helo Blod sy’n gasgliad o adnoddau defnyddiol i ddefnyddwyr a datblygwyr. Yn fewnol, rydym yn cydweithio’n barhaus gyda swyddogion y Gwasanaeth Cyfieithu, Safonau a’r Swyddfa AI i godi ymwybyddiaeth o’r ffyrdd y gall technoleg helpu staff i weithio’n fwy dwyieithog.

[redacted information]

[WG Official 1]

ATISN 27572 - I nodi ogydd: FOI ATISM 27499

From: [WG Official 1]

Sent: 24 September 2026 09:40

To: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]

Cc: [WG Official 2]; [WG Official 3]; [WG Official 4]; [Freedom of Information Officer]; [PS First Minister Wales/YP Prif Weinidog Cymru]; [Welsh Government Newsdesk]; [WG Official 5]; [LCC - Freedom of Information / Rhyddid Gwybodaeth]

Subject: ATB/RE: I nodi ogydd: FOI ATISM 27499

Annwyl Swyddfa Breifat,

Diolch am y sgwrs.

Dyma gadarnhau nad oes gofyn i'r Gweinidog glirio'r ymateb, dim ond i nodi. Byddwn ni'n danfon yr ymateb erbyn y dyddiad terfynol isod.

Diolch,

[WG Official 1]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Mercher, 23 Medi 2026 16:34

At: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet

dros Addysg a'r Gymraeg]

Copi/Cc: [WG Official 2]; [WG Official 3]; [WG Official 4]; [FOI / Rhyddid Gwybodaeth - ETCC]; [PS First Minister Wales/YP Prif Weinidog Cymru]; [Welsh Government Newsdesk]; [WG Official 5]

Pwnc: I nodi ogydd: FOI ATISM 27499

Annwyl Swyddfa Breifat,

Dyma ymateb i FOI ATISN 27499 yn ymwneud â manylion partneriaeth Llywodraeth Cymru a Cosine.

A fyddai'r Gweinidog yn fodlon nodi yr ymateb os gwelwch yn dda, a chadarnhau hynny er mwyn i ni ymateb i'r ymholydd erbyn **dydd Gwener 02 Hydref**.

Diolch,

[WG Official 1]

ATISN 27572 - Email - 20260910 - "RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept"

Oddi wrth: [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Anfonwyd: Dydd Iau, 10 Medi 2026 16:03

At: [WG Official 1]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Copi/Cc: [WG Official 2]; [Education Welsh Language Press Office]; [WG Official 3]; [WG Official 4]; [WG Official 5]; [WG Official 6]; [WG Official 7]; [WG Official 8]; [Economy Press Office]; [WG Official 9]; [WG Official 10]; [WG Official 11]; [WG Official 12]; [WG Official 13]; [WG Official 14]; [WG Official 15]; [WG Official 16]; [WG Official 17]; [WG Official 18]; [WG Official 19]; [WG Official 20]; [WG Official 21]; [WG Official

22]; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]

Pwnc: RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Hi [WG Official 1]

The CMECE is content to clear this press notice.

Apologies for the delay in getting back to you

Regards

[WG Official 24]

From: [WG Official 1]

Sent: 07 September 2026 16:49

To: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Cc: [WG Official 2]; [Education Welsh Language Press Office]; [WG Official 3]; [WG Official 4]; [WG Official 5]; [WG Official 6]; [WG Official 7]; [WG Official 8]; [Economy Press Office]; [WG Official 9]; [WG Official 10]; [WG Official 11]; [WG Official 12]; [WG Official 13]; [WG Official 14]; [WG Official 15]; [WG Official 16]; [WG Official 17]; [WG Official 18]; [WG Official 19]; [WG Official 20]; [WG Official 21]; [WG Official 22]; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]

Subject: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Hello

Please find attached a draft PN on a new partnership with AI company Cosine to advance Welsh language AI.

Can we get clearance by both CMs by 5pm on Wednesday please?

Diolch yn fawr

[WG Official 1]

ATISN 27572 - Email - 20260909 - "ATB/RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept" REDACTED

Oddi wrth: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]

Anfonwyd: Dydd Mercher, 9 Medi 2026 13:07

At: [WG Official 1]; [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Copi/Cc: [WG Official 2]; [Education Welsh Language Press Office]; [WG Official 3]; [WG Official 4]; [WG Official 5]; [WG Official 6]; [WG Official 7]; [WG Official 8]; [Economy Press Office]; [WG Official 9]; [WG Official 10]; [WG Official 11]; [WG Official 12]; [WG Official 13]; [WG Official 14]; [WG Official 15]; [WG Official 16]; [WG Official 17]; [WG Official 18]; [WG Official 19]; [WG Official 20]; [WG Official 21]; [WG Official 22]; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]

Pwnc: ATB/RE: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Prynhawn da [WG Official 1],

Mae'r Gweinidog Cabinet wedi clirio / cmewl has cleared subject to changing "The Cabinet Minister for Education and the Welsh Language said" to say "[WG Official 24] said" please.

Diolch,

[WG Official 25]

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Llun, 7 Medi 2026 16:49

At: [PS Cabinet Minister for Education and the Welsh Language/YP Gweinidog Cabinet dros Addysg a'r Gymraeg]; [PS Cabinet Minister for Enterprise, Connectivity and Energy/YP Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni]

Copi/Cc: [WG Official 2]; [Education Welsh Language Press Office]; [WG Official 3]; [WG Official 4]; [WG Official 5]; [WG Official 6]; [WG Official 7]; [WG Official 8]; [Economy Press Office]; [WG Official 9]; [WG Official 10]; [WG Official 11]; [WG Official 12]; [WG Official 13]; [WG Official 14]; [WG Official 15]; [WG Official 16]; [WG Official 17]; [WG Official 18]; [WG Official 19]; [WG Official 20]; [WG Official 21]; [WG Official 22]; [Elin Burns {then} Director Culture, Heritage, Sport & Welsh language]

Pwnc: For CMEWL and CMECE to clear by 5pm Wednesday 9th Sept

Hello

Please find attached a draft PN on a new partnership with AI company Cosine to advance Welsh language AI.

Can we get clearance by both CMs by 5pm on Wednesday please?

Diolch yn fawr

[WG Official 1]

ATISN 27572 – Email – 20260924 - TO/AB/05874/26 ATB/RE: Urgent concerns regarding Welsh Government collaboration with Cosine on "Lumen Sovereign" model

Oddi wrth: [WG Official 1]

Anfonwyd: Dydd Iau, 24 Medi 2026 10:59

At: [Person 1]

Copi/Cc: [Correspondence/Gohebiaeth – AB]; [Correspondence/Gohebiaeth – AP]; [Cymraeg 2050]

Pwnc: TO/AB/05874/26 ATB/RE: Urgent concerns regarding Welsh Government collaboration with Cosine on "Lumen Sovereign" model

Annwyl [Person 1]

Thank you for your messages to the Cabinet Minister for Education and the Welsh Language, and the Cabinet Minister for Enterprise, Connectivity and Energy. I have been asked to respond.

You raise several questions regarding the Welsh Government's collaboration with Cosine. I have addressed them under the themes as set out in your correspondence.

1. Licensing and Open Weights / Data Governance and Commercial Value

Our partnership with Cosine is based on sharing advice and open, publicly available resources. We have only signposted to high-quality open data sources which are already publicly available, will and remain available, for anyone to use and reuse. We have not transferred ownership of any datasets or resources to Cosine.

We do not determine how Cosine, or any other organisation, may reuse those open, publicly available resources. We do not therefore hold information on any future model weights, licensing arrangements, training pipelines, pricing models or commercial products that may be developed by third parties.

2. Protection Against Enclosure / Open Source vs Vendor Lock-In

We believe that Cymraeg belongs to us all and that there should be no barriers for anyone to use our language—individuals, or corporations. Through our Welsh language

technology grants, we have invested—and continue to invest—in Welsh-language technology and AI, e.g. via our annual grant to Bangor University (which has recently received a 3-year funding offer), funding a range of tools, data and resources that are publicly available free of charge for anyone to use and reuse.

The resources referred to in the partnership are open resources which were already publicly available and remain available to any individual or organisation under their existing terms and licences. We have not provided exclusive access to these or any resources.

3. Support for Domestic Research / Strategic Technology Infrastructure in Wales

The Welsh Government's policy to provide advice on Welsh-language technology and AI, and to signpost organisations to open, publicly available resources applies to any organisation seeking advice in this area. Following an approach from Cosine seeking advice on Welsh-language AI, we formed a partnership based on sharing expertise and signposting relevant publicly available open resources.

We have invested, and continue to invest, in Welsh universities and open-source, publicly available natural language processing components through our Welsh-language technology grant funding. This funding has supported the development of Welsh-language tools, datasets and resources and most recently funded Bangor University's Iriaith project, Cardiff University's NLP Cymru CyTag infrastructure, Data Orchard's Mapio Cymru project, and the National Library of Wales' Wici Cymru a'i Phobl project. All Welsh-language technology resources funded by the Welsh Government are made available openly and free of charge for anyone to use and reuse.

Yn gywir,

[WG Official 1]

Oddi wrth: [Person 1]

Anfonwyd: Dydd Gwener, 11 Medi 2026 12:59

At: [Correspondence/Gohebiaeth – AB]

Pwnc: Urgent concerns regarding Welsh Government collaboration with Cosine on "Lumen Sovereign" model

Annwyl Weinidog,

I am writing to you as a Welsh-language technologist and open-source software practitioner to express serious concern regarding the Welsh Government's announced collaboration with Cosine, an external commercial entity, to feed Welsh-language data into the training pipeline for their proprietary "Lumen Sovereign" model.

Welsh-language digital text and speech data represent decades of public investment, community effort, and national heritage. Under the Welsh Language Technology Action Plan, public policy has rightly recognized that low-resource languages require deliberate infrastructure to thrive digitally. However, true linguistic sovereignty cannot be achieved by transferring publicly gathered language resources to an external, private company to develop a closed commercial product.

I urge you to clarify the following:

1. Licensing and Open Weights: Has the Welsh Government attached reciprocal open-source conditions to this data sharing? Will the resulting model weights, training pipeline recipes, and curated corpora be released under permissive open-source licenses (e.g., Apache 2.0, MIT) for the benefit of all Welsh developers, schools, and public bodies?
2. Protection Against Enclosure: What guarantees are in place to ensure that the Welsh language is not enclosed behind proprietary API paywalls, requiring public sector bodies and Welsh citizens to purchase access to capabilities built using their own cultural data?
3. Support for Domestic Research: Why was this strategic initiative prioritized over providing equivalent compute, data access, and procurement support to indigenous Welsh NLP projects and university research initiatives (such as Bangor University's Canolfan Bedwyr or Cardiff University's NLP Cymru)?

A language's digital future must not be treated as raw material for external private equity. I look forward to your response on how the Welsh Government intends to safeguard our linguistic commons.

Yn gywir,

[Person 1]

Attachments

ATISN 27572 – Email – 20260813 – “YML/FW: ATB/RE: Potential Cosine partnership announcement” - attachment

Cosine and Llywodraeth Cymru (Welsh Government) collaborate to support Welsh language representation in Lumen model training

The collaboration will bring Welsh language data into Cosine’s Lumen training pipeline as part of a broader effort to build AI systems that better represent the UK.

[London/Cardiff], [Date] — Cosine today announced a collaboration with Llywodraeth Cymru (Welsh Government) to support the inclusion of Welsh language data in training for its Lumen Sovereign model.

The work will help ensure Lumen Sovereign reflects the UK more accurately in practice. That includes Welsh, one of the UK’s living national languages and an important part of public, cultural, and institutional life in Wales.

As AI systems become part of everyday software, services, and digital infrastructure, language coverage matters. Models used across the UK should represent the languages and contexts that shape life here. This collaboration with the Welsh Government brings Welsh into that work in a direct and practical way and is part of a broader effort to build models that are more useful, better informed, and more representative of the country as a whole.

Alistair Pullen, CEO, Cosine

“Welsh is an essential part of the UK’s linguistic landscape. This work with the Welsh Government helps us build Lumen on a stronger foundation, with better representation of the languages and public context that matter across the UK.”

[Welsh Government spokesperson quote placeholder]

“The Welsh language is a vital part of Wales’s identity and public life. Supporting its inclusion in AI model development is an important step in ensuring new technologies better represent the people and languages of the UK.”

This collaboration sits within a wider effort to make Lumen more representative of the UK, including the languages, knowledge, and public context that general-purpose model development often overlooks.

Further details on the collaboration will be shared in due course.

About Cosine

Cosine is the British sovereign AI lab building frontier models for defence, national

security, and highly regulated industries. Our Lumen model family is designed for software engineering and other high-value tool-using work in fully air-gapped, high-security environments.

Powered by our in-house ML research lab and backed by the UK Government's Sovereign AI initiative, we are building the UK's first fully sovereign frontier AI model. While we also post-train strong open-weight models on proprietary human-reasoning and domain-specific data, our flagship sovereign model is being developed with zero dependence on foreign infrastructure. This gives customers frontier capability with stronger control over training, deployment, and long-context performance.

Deployable entirely within a customer's VPC or fully on-premise with zero data egress, Cosine gives governments, enterprises, and critical infrastructure providers sovereign control over the models, infrastructure, and data behind their AI systems.

ATISN 27572 - Briefing - 20260909 - "Tech briefing with CMEWL (CY)"

Technoleg a deallusrwydd artiffisial (AI) Cymraeg

Pwrpas y cyfarfod

Rhwng mis Medi a Rhagfyr 2026, rydym yn cynllunio cyfres o gyhoeddiadau yn ymwneud â thechnoleg ac AI Cymraeg. Pwrpas y cyfarfod yw briffio y Gweinidog Cabinet ar y cefndir a'r datblygiadau diweddaraf yn y maes.

Cefndir

Gall technoleg Cymraeg cywir, dibynadwy ein helpu ni i gynyddu'r defnydd dyddiol o'r Gymraeg, sy'n cyfrannu'n uniongyrchol at brif dargedau Cymraeg 2050. Gall technoleg ei gwneud hi'n haws i ni ddefnyddio'r Gymraeg ym mhob agwedd o fywyd, o addysg a gwaith i sgwrsio gyda ffrindiau, teulu a chydweithwyr.

O dan y Llywodraeth flaenorol, lluniwyd ein blaenoriaethau ar gyfer technoleg ac AI Cymraeg, sef:

- technoleg fel ffordd o gynyddu'r defnydd dyddiol o'r Gymraeg
- sicrhau bod pawb yn gallu cael at dechnoleg Cymraeg
- gwella deallusrwydd artiffisial (AI) Cymraeg a thechnolegau lleferydd (drwy rannu data a ffyrdd eraill)

Fel rhan o'r gwaith hwn, rydyn ni wedi cydweithio mewn partneriaeth gyda Microsoft i greu'r cyfleuster cyfieithu ar y pryd dynol yn Teams. Hwn oedd y cyntaf o'i fath i Microsoft ac mae bellach yn cael ei ddefnyddio o amgylch y byd, a hyn i gyd wedi'i seilio ar ein gwaith ni yma yng Nghymru.

Bellach, mae Microsoft wedi cyhoeddi bod Copilot yn gweithio yn Gymraeg, a bod capsiynau Teams yn amlieithog hefyd—a'r gallu i'w cyfieithu ar y pryd. Mae hyn yn adlewyrchu gwaith pwysig rhyngom ni a Microsoft i gryfhau sut mae technoleg yn cefnogi defnyddwyr amlieithog sydd am ddefnyddio mwy nag un iaith ar yr un pryd. Y bwriad yw parhau i gydweithio mewn partneriaeth gyda Microsoft ar achos busnes i wella sut mae eu hadnoddau yn gweithio yn Gymraeg ac yn ddwyieithog, yng Nghymru ac yn rhyngwladol.

Bwydodd hyn oll fewn i beilot mewnol "*Gweithio Bilingually*" i ystyried sut all dechnoleg gefnogi ni i weithio'n ddwyieithog. Dangosodd y peilot fod y technolegau hyn yn gallu cynyddu hyder a defnydd o'r Gymraeg, ac rydym bellach yn bwrw ymlaen â'r argymhellion i ehangu'r defnydd ohonynt ar draws y sefydliad.

Cyhoeddiadau

Partneriaeth rhwng Cosine a Llywodraeth Cymru (disgwyl wythnos sy'n cychwyn 14 Medi 2026)

Mae Llywodraeth Cymru yn bwriadu cyhoeddi partneriaeth â chwmni AI sofran Cosine i helpu gwella sut mae AI yn gweithio yn Gymraeg. Bydd hwn yn cyd-fynd ag uchelgeisiau ehangach y DU i gryfhau gallu AI sofran, gan helpu i sicrhau bod y Gymraeg yn cael ei chynrychioli mewn systemau AI sy'n cael eu datblygu yn y DU. Bydd y bartneriaeth yn adeiladu ar ein dull hirsefydlog o rannu data, adnoddau ac arbenigedd er mwyn gwella sut mae AI yn prosesu'r Gymraeg. Dylid nodi nad oes ymrwymiad ariannol i Lywodraeth Cymru; rhannu arbenigedd ac adnoddau agored fydd sail y bartneriaeth hon.

Bydd swyddogion perthnasol hefyd yn briffio'r Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni a'r Gweinidog Cabinet dros Effeithiolrwydd Llywodraeth a'r Cyfansoddiad, gan adlewyrchu'r cysylltiadau gydag AI, technoleg a thrawsnewid gwasanaethau cyhoeddus. Bydd swyddfeydd perthnasol felly yn cydgysylltu ar unrhyw gyhoeddiadau.

[redacted information]

Er eich gwybodaeth: yn 2024 cyhoeddodd Llywodraeth Cymru ar y pryd bartneriaeth gyda chwmni AI OpenAI i wella perfformiad modelau AI ChatGPT yn Gymraeg.

UNESCO (disgwyl 30 Medi 2026)

Fel rhan o Ddegawd Rhyngwladol leithoedd Brodorol UNESCO, rydym wedi gweithio gydag UNESCO i addasu ein pecyn cymorth ar ddatblygu meddalwedd a gwasanaethau dwyieithog, ar gyfer cynulleidfa ryngwladol.

Bydd y fersiwn newydd yn rhannu arferion da o Gymru gyda chymunedau rhyngwladol, i helpu ieithoedd llai eu hadnoddau i ddatblygu gwasanaethau a meddalwedd dwyieithog. Mae'r gwaith yn enghraifft o Gymru yn cael ei chydabod ar lefel rhyngwladol fel arweinydd yn y maes. UNESCO fydd yn cyhoeddi'r pecyn gyda chydabyddiaeth i Lywodraeth Cymru. Mae'r Gweinidog Cabinet wedi cytuno i gynnwys y dyfyniad yma yn y rhagair:

“As Cabinet Minister for Education and the Welsh Language, I’m pleased to see our commitment to linguistic diversity and digital inclusion being recognised beyond our borders. By sharing what we’ve learned from strengthening Welsh in our own digital spaces, we’re contributing to a wider international conversation, and helping other multilingual communities take similar steps. Together, we’re proudly working with our friends and partners around the world to ensure that the digital world supports and reflects our languages.”

Er eich gwybodaeth: aeth swyddogion gyda'r Prif Weinidog ar y pryd i Baris yn 2025 i drafod cydweithio gydag UNESCO ar dechnoleg iaith a hwn oedd un o'r pwyntiau gweithredu.

Ymgysylltu â chwmnïau technoleg

Ymrwymodd y Gweinidog Cabinet i ddanfôn llythyrau at gwmnïau “big tech” yn ei datganiad llafar ar 07 Gorffennaf 2026, a hyn i sbarduno trafodaethau am gefnogaeth i'r Gymraeg mewn technoleg ac AI.

Wrth i ni dderbyn ymatebion, byddwn yn edrych i drefnu cyfarfodydd rhwng y Gweinidog Cabinet a'r cwmnïau hynny, os bydd y Gweinidog Cabinet yn cytuno. Bydd cyngor pellach yn cael ei ddarparu wrth i ymatebion gael eu derbyn.

[redacted information]

Ein gwaith ar waith

Rydym wedi creu adran newydd ar dudalen Technoleg Cymraeg Llywodraeth Cymru i rannu enghreifftiau ymarferol o sut mae pobl yn defnyddio'n adnoddau. Dros y misoedd nesaf byddwn yn defnyddio'r adran hon a'n sianeli cyfathrebu i rannu rhagor o

enghreiffiau ac i godi ymwybyddiaeth o'r hyn sydd ar gael. Ymhlith yr enghreifftiau cyntaf rydym wedi'u cyhoeddi mae:

Comisiynydd y Gymraeg yn defnyddio trawsgrifio Teams a Copilot Cymraeg i greu crynodebau a chymau gweithredu o gyfarfodydd Cymraeg a dwyieithog;

Swyddfa'r Wasg Llywodraeth Cymru yn defnyddio Trawsgrifiwr (o brosiect grant Prifysgol Bangor) i gofnodi cyfweiliadau Cymraeg;

Senedd Cymru yn defnyddio data o brosiect grant Mapio Cymru fel sail i'w map etholaethau rhyngweithiol.

ATISN 27572 - Briefing - 20260922 - "Technoleg Cymraeg - partneriaeth Cosine"

Fersiwn dydd Mawrth 22/9/26

Technoleg Cymraeg - Partneriaeth Cosine

Llinellau

Ry'n ni eisiau i'r Gymraeg fod yn iaith mae modd ei defnyddio yn y dechnoleg mae pobl yn eu defnyddio fwyfwy—yn y gwaith, mewn addysg, mewn gwasanaethau cyhoeddus ac ym mywyd bob dydd. Fwyfwy, mae pobl yn defnyddio AI. Ac mae'n bwysig bod gan y Gymraeg le yn y byd hwn sy'n datblygu mor gyflym.

Dyna pam ry'n ni'n ariannu cydrannau technoleg Cymraeg ac yn eu gwneud ar gael yn gyhoeddus ac am ddim—i Cosine, ac i *unrhyw un* eu defnyddio a'u hailddefnyddio.

Ry'n ni'n gwneud hynny fel bod unrhyw un—datblygwyr, ymchwilwyr, busnesau, cyrff cyhoeddus, grwpiau cymunedol neu gwmnïau technoleg—yn gallu defnyddio ac ailddefnyddio'r adnoddau hynny i gryfhau'r Gymraeg.

Fel Gweinidog, dw i eisoes wedi ysgrifennu at nifer o brif gwmnïau technoleg am eu darpariaeth Gymraeg, ac dw i'n edrych ymlaen at eu cyfarfod [*If pressed: rydych wedi sgwennu at Microsoft, Apple, Amazon, Google, Meta, OpenAI, Anthropic.*

Mae pawb wedi ymateb heblaw am Google. Mae cyfarfodydd wedi'u trefnu at ddiwedd y flwyddyn].

Os yw cwmnïau technoleg yn barod i drafod sut y gallan nhw wella darpariaeth Gymraeg, ry'n ni'n barod i drafod gyda nhw.

Mae'r diben yn syml: sicrhau bod gan y Gymraeg le cryf yn y byd digidol.

Pam gweithio gyda Cosine?

Daeth Cosine at Lywodraeth Cymru i geisio cyngor ac arbenigedd ym maes AI Cymraeg.

Mewn ymateb, ffurfion ni bartneriaeth sy'n seiliedig ar rannu arbenigedd a chyfeirio at adnoddau Cymraeg sydd eisoes ar gael yn gyhoeddus.

Does dim statws arbennig nac unigryw i Cosine o dan y trefniant hwn.

Yr hyn wnaethon ni gyda Cosine yw'r hyn fydden ni'n ei wneud gydag unrhyw un sy'n dod aton ni am gyngor yn y maes hwn: rhannu arbenigedd, a'u cyfeirio at adnoddau Cymraeg agored sydd eisoes ar gael yn gyhoeddus i bawb.

Felly, dyw'r bartneriaeth ddim yn ymwneud â ffafrio un cwmni. Mae'n ymwneud â sicrhau bod sefydliadau sydd eisiau gwella darpariaeth Gymraeg yn gallu cael eu cyfeirio at adnoddau sydd eisoes ar gael i bawb. Ry'n ni'n awyddus i helpu unrhyw un sy'n ymddiddori mewn technoleg a'r Gymraeg.

Faint o arian ry'n ni'n ei dalu i Cosine?

Dim ceiniog.

Dyw hwn ddim yn grant.

Dyw hwn ddim yn gontract.

Dyw hwn ddim yn ymarfer caffael.

Does dim arian yn cael ei dalu i Cosine drwy'r trefniant hwn.

Mae'r bartneriaeth yn ymwneud â rhannu arbenigedd a chyfeirio at adnoddau agored sydd eisoes ar gael yn gyhoeddus.

Beth sydd nesaf?

Byddwn ni'n parhau i gynghori unrhyw un sydd am wella'r Gymraeg ym myd technoleg, a byddwn ni'n parhau i ddatblygu offer a chydannau technoleg agored i bobl ddefnyddio ac ailddefnyddio yn eu cynnyrch. A byddwn ni'n sicrhau bod rhain ar gael yn rhad ac am ddim.

If pressed: Fel cam nesaf yn y bartneriaeth, mae Cosine eisiau cynnal trafodaeth bord gron gyda Llywodraeth Cymru ar ddefnyddio AI yn y sector cyhoeddus. Byddai hyn yn adlewyrchu ffocws Cynllun AI i Gymru ar gyflawni, gan roi'r trefniadau llywodraethu, y gallu a'r cymorth ymarferol sydd eu hangen i helpu sefydliadau i fabwysiadu AI yn ddiogel, yn foesebol, yn effeithiol ac yn ddwyieithog. Bydd swyddogion yn briffio Gweinidogion ar hyn maes o law.

Pa ddata ry'n ni'n ei ryddhau i Cosine?

Mae'r bartneriaeth yn seiliedig ar rannu arbenigedd a chyfeirio at adnoddau agored sydd eisoes ar gael yn gyhoeddus, ac sy'n parhau i fod ar gael i unrhyw un eu defnyddio a'u hailddefnyddio.

Dyw Llywodraeth Cymru ddim wedi trosglwyddo perchnogaeth unrhyw setiau data nac adnoddau iaith i Cosine.

Mae hyn yn golygu nad yw Cosine yn cael mynediad arbennig at ddata preifat, cyfrinachol neu sensitif Llywodraeth Cymru.

Mae'r adnoddau hyn yn cynnwys ffynonellau data agored o safon uchel, gan gynnwys adnoddau a ddatblygwyd drwy ymchwil academiaidd a gweithgareddau a ariannwyd gan Lywodraeth Cymru (er enghraifft, gan Brifysgol Bangor).

Gan fod yr adnoddau hyn yn agored ac eisoes ar gael yn gyhoeddus, maen nhw, wrth eu natur, ar gael i unrhyw un eu defnyddio a'u hailddefnyddio.

Llinellau cyffredinol

Mae'r Gymraeg yn perthyn i ni i gyd. Mae'n rhan o bwy ydyn ni fel cenedl, ac rydyn ni am greu mwy o gyfleoedd i ni gyd ei defnyddio ym mhob rhan o'n bywydau.

Mae'n bwysig bod y Gymraeg yn cael ei chynrychioli'n llawn yn y technolegau y mae pobl yn eu defnyddio i gael gafael ar wybodaeth, dysgu, gweithio a chyfathrebu.

Mae'r bartneriaeth newydd gyda Cosine yn helpu i gryfhau gallu Cymraeg mewn AI, gan greu mwy o gyfleoedd i ni i gyd ddefnyddio ein Cymraeg yn hyderus bob dydd a chefnogi dyfodol yr iaith yn yr oes ddigidol.

Llinellau amddiffynol

Gadewch imi fod yn glir ynghylch beth yw'r bartneriaeth hon—a beth nad yw hi.

Does dim arian yn newid dwylo.

Does dim ymrwymiad ariannol gan Lywodraeth Cymru.

Does dim data preifat na sensitif na chyfrinachol yn cael ei rannu gan Lywodraeth Cymru.

Mae'r deunydd dan sylw eisoes ar gael yn gyhoeddus ac yn agored.

Dyw hwn ddim yn drefniant unigryw. Bydden ni'n croesawu sgysrsiau gyda chwmnïau technoleg o bob math. Dyna pam dwi eisoes wedi ysgrifennu at rai o'r prif rai ac yn edrych ymlaen at eu cyfarfod.

Mae'n gam ymarferol i helpu i wella'r Gymraeg mewn technolegau sy'n datblygu.

Mae'r ffeithiau'n syml: dim arian, dim data preifat, dim trefniant unigryw, ac adnoddau sydd eisoes ar gael yn gyhoeddus ac am ddim.

Drwy grantiau technoleg, mae Llywodraeth Cymru wedi buddsoddi, ac yn parhau i fuddsoddi, mewn technoleg ac AI Cymraeg, e.e. drwy Brifysgol Bangor. Mae hyn yn cynnwys ariannu offer, data ac adnoddau sydd ar gael yn rhad ac am ddim i unrhyw un eu defnyddio a'u hailddefnyddio.

AI moesegol

Dw i'n gwybod bod lot wedi bod yn y cyfryngau yn ddiweddar am AI. Ry'n ni'n monitro ac yn adolygu'r sefyllfa yn barhaus.

Ry'n ni wedi ymrwymo i sicrhau bod AI yn cael ei fabwysiadu mewn ffyrdd diogel, moesegol a thryloyw, gyda goruchwyliaeth ddynol briodol a chyfrifoldeb clir am wneud penderfyniadau.

Ry'n ni'n datblygu fframwaith llywodraethu gweithredol cenedlaethol ar gyfer AI i helpu gwasanaethau cyhoeddus Cymru i fabwysiadu AI yn gyfrifol.

Cefndir

Mae deallusrwydd artiffisial (AI) yn datblygu'n gyflym ledled y byd, ac yn cynnig cyfleoedd sylweddol ond hefyd godi cwestiynau am ddiogelwch, preifatrwydd a rheoleiddio. Yn ddiweddar, mae nifer o arbenigwyr a llunwyr polisi rhyngwladol wedi galw am fwy o oruchwyliaeth er mwyn sicrhau bod AI yn cael ei ddatblygu a'i ddefnyddio mewn ffordd ddiogel, foesegol ac atebol.

Yn erbyn y cefndir hwn, cyhoeddodd Llywodraeth Cymru a Cosine eu bod yn cydweithio i gryfhau deallusrwydd artiffisial Cymraeg ym mis Medi 2026. Mae'r bartneriaeth yn seiliedig ar rannu arbenigedd a chyfeirio at adnoddau Cymraeg agored sydd eisoes ar gael yn gyhoeddus yn rhad ac am ddim. Does dim ymrwymiad ariannol gan Lywodraeth Cymru o dan y trefniant hwn, a does dim trosglwyddo perchnogaeth unrhyw setiau data nac adnoddau iaith i Cosine.

Mae'r gwaith hwn yn adeiladu ar fuddsoddiad parhaus Llywodraeth Cymru mewn technoleg Gymraeg e.e. drwy Brifysgol Bangor, gan gynnwys partneriaeth flaenorol gydag OpenAI yn 2024.

Swyddog: [WG Official 1] ac [WG Official 2] 22/09/2026)

ATISN 27572 - Briefing - 20260922 - "YML/FW: Leins Ychwanegol Cosine"

Fersiwn dydd Mawrth 22/9/26

Technoleg Gymraeg - Partneriaeth Cosine

Llinellau

Ry'n ni eisiau i'r Gymraeg fod yn iaith mae modd ei defnyddio yn y dechnoleg mae pobl yn eu defnyddio fwyfwy—yn y gwaith, mewn addysg, mewn gwasanaethau cyhoeddus ac ym mywyd bob dydd.

Fwyfwy, mae pobl yn defnyddio AI. Ac mae'n bwysig bod gan y Gymraeg le yn y byd hwn sy'n datblygu mor gyflym.

Dyna pam ry'n ni'n ariannu cydrannau technoleg Cymraeg ac yn eu gwneud ar gael yn gyhoeddus ac am ddim—i Cosine, ac i *unrhyw un* eu defnyddio a'u hailddefnyddio.

Ry'n ni'n gwneud hynny fel bod unrhyw un—datblygwyr, ymchwilwyr, busnesau, cyrff cyhoeddus, grwpiau cymunedol neu gwmnïau technoleg—yn gallu defnyddio ac ailddefnyddio'r adnoddau hynny i gryfhau'r Gymraeg.

Fel Gweinidog, dw i eisoes wedi ysgrifennu at nifer o brif gwmnïau technoleg am eu darpariaeth Gymraeg, ac dw i'n edrych ymlaen at eu cyfarfod [*If pressed: rydych wedi sgwennu at Microsoft, Apple, Amazon, Google, Meta, OpenAI, Anthropic. Mae pawb wedi ymateb heblaw am Google. Mae cyfarfodydd wedi'u trefnu at ddiwedd y flwyddyn*].

Os yw cwmnïau technoleg yn barod i drafod sut y gallan nhw wella darpariaeth Gymraeg, ry'n ni'n barod i drafod gyda nhw.

Mae'r diben yn syml: sicrhau bod gan y Gymraeg le cryf yn y byd digidol.

Pam gweithio gyda Cosine?

Daeth Cosine at Lywodraeth Cymru i geisio cyngor ac arbenigedd ym maes AI Cymraeg.

Mewn ymateb, ffurfion ni bartneriaeth sy'n seiliedig ar rannu arbenigedd a chyfeirio at adnoddau Cymraeg sydd eisoes ar gael yn gyhoeddus.

Does dim statws arbennig nac unigryw i Cosine o dan y trefniant hwn.

Yr hyn wnaethon ni gyda Cosine yw'r hyn fydden ni'n ei wneud gydag unrhyw un sy'n dod aton ni am gyngor yn y maes hwn: rhannu arbenigedd, a'u cyfeirio at adnoddau Cymraeg agored sydd eisoes ar gael yn gyhoeddus i bawb.

Felly, dyw'r bartneriaeth ddim yn ymwneud â ffafrio un cwmni. Mae'n ymwneud â sicrhau bod sefydliadau sydd eisiau gwella darpariaeth Gymraeg yn gallu cael eu cyfeirio at adnoddau sydd eisoes ar gael i bawb. Ry'n ni'n awyddus i helpu unrhyw un sy'n ymddiddori mewn technoleg a'r Gymraeg.

Faint o arian ry'n ni'n ei dalu i Cosine?

Dim ceiniog.

Dyw hwn ddim yn grant.

Dyw hwn ddim yn gontract.

Dyw hwn ddim yn ymarfer caffael.

Does dim arian yn cael ei dalu i Cosine drwy'r trefniant hwn.

Mae'r bartneriaeth yn ymwneud â rhannu arbenigedd a chyfeirio at adnoddau agored sydd eisoes ar gael yn gyhoeddus.

Beth sydd nesaf?

Byddwn ni'n parhau i gynghori unrhyw un sydd am wella'r Gymraeg ym myd technoleg, a byddwn ni'n parhau i ddatblygu offer a chydrannau technoleg agored i bobl ddefnyddio ac ailddefnyddio yn eu cynnyrch. A byddwn ni'n sicrhau bod rhain ar gael yn rhad ac am ddim.

If pressed: Fel cam nesaf yn y bartneriaeth, mae Cosine eisiau cynnal trafodaeth bord gron gyda Llywodraeth Cymru ar ddefnyddio AI yn y sector cyhoeddus. Byddai hyn yn adlewyrchu ffocws Cynllun AI i Gymru ar gyflawni, gan roi'r trefniadau llywodraethu, y gallu a'r cymorth ymarferol sydd eu hangen i helpu sefydliadau i fabwysiadu AI yn ddiogel, yn foesegol, yn effeithiol ac yn ddwyieithog. Bydd swyddogion yn briffio Gweinidogion ar hyn maes o law.

Pa ddata ry'n ni'n ei ryddhau i Cosine?

Mae'r bartneriaeth yn seiliedig ar rannu arbenigedd a chyfeirio at adnoddau agored sydd eisoes ar gael yn gyhoeddus, ac sy'n parhau i fod ar gael i unrhyw un eu defnyddio a'u hailddefnyddio.

Dyw Llywodraeth Cymru ddim wedi trosglwyddo perchnogaeth unrhyw setiau data nac adnoddau iaith i Cosine.

Mae hyn yn golygu nad yw Cosine yn cael mynediad arbennig at ddata preifat, cyfrinachol neu sensitif Llywodraeth Cymru.

Mae'r adnoddau hyn yn cynnwys ffynonellau data agored o safon uchel, gan gynnwys adnoddau a ddatblygwyd drwy ymchwil academiaidd a gweithgareddau a ariannwyd gan Lywodraeth Cymru (er enghraifft, gan Brifysgol Bangor).

Gan fod yr adnoddau hyn yn agored ac eisoes ar gael yn gyhoeddus, maen nhw, wrth eu natur, ar gael i unrhyw un eu defnyddio a'u hailddefnyddio.

Llinellau cyffredinol

Mae'r Gymraeg yn perthyn i ni i gyd. Mae'n rhan o bwy ydyn ni fel cenedl, ac rydyn ni am greu mwy o gyfleoedd i ni gyd ei defnyddio ym mhob rhan o'n bywydau.

Mae'n bwysig bod y Gymraeg yn cael ei chynrychioli'n llawn yn y technolegau y mae pobl yn eu defnyddio i gael gafael ar wybodaeth, dysgu, gweithio a chyfathrebu.

Mae'r bartneriaeth newydd gyda Cosine yn helpu i gryfhau gallu Cymraeg mewn AI, gan greu mwy o gyfleoedd i ni i gyd ddefnyddio ein Cymraeg yn hyderus bob dydd a chefnogi dyfodol yr iaith yn yr oes ddigidol.

Llinellau amddiffynol

Gadewch imi fod yn glir ynghylch beth yw'r bartneriaeth hon—a beth nad yw hi.

Does dim arian yn newid dwylo.

Does dim ymrwymiad ariannol gan Lywodraeth Cymru.

Does dim data preifat na sensitif na chyfrinachol yn cael ei rannu gan Lywodraeth Cymru.

Mae'r deunydd dan sylw eisoes ar gael yn gyhoeddus ac yn agored.

Dyw hwn ddim yn drefniant unigryw. Bydden ni'n croesawu sgysiau gyda chwmnïau technoleg o bob math. Dyna pam dwi eisoes wedi ysgrifennu at rai o'r prif rai ac yn edrych ymlaen at eu cyfarfod.

Mae'n gam ymarferol i helpu i wella'r Gymraeg mewn technolegau sy'n datblygu.

Mae'r ffeithiau'n syml: dim arian, dim data preifat, dim trefniant unigryw, ac adnoddau sydd eisoes ar gael yn gyhoeddus ac am ddim.

Drwy grantiau technoleg, mae Llywodraeth Cymru wedi buddsoddi, ac yn parhau i fuddsoddi, mewn technoleg ac AI Cymraeg, e.e. drwy Brifysgol Bangor. Mae hyn yn cynnwys ariannu offer, data ac adnoddau sydd ar gael yn rhad ac am ddim i unrhyw un eu defnyddio a'u hailddefnyddio.

AI moesegol

Dw i'n gwybod bod lot wedi bod yn y cyfryngau yn ddiweddar am AI. Ry'n ni'n monitro ac yn adolygu'r sefyllfa yn barhaus.

Ry'n ni wedi ymrwymo i sicrhau bod AI yn cael ei fabwysiadu mewn ffyrdd diogel, moesegol a thryloyw, gyda goruchwyliaeth ddynol briodol a chyfrifoldeb clir am wneud penderfyniadau.

Ry'n ni'n datblygu fframwaith llywodraethu gweithredol cenedlaethol ar gyfer AI i helpu gwasanaethau cyhoeddus Cymru i fabwysiadu AI yn gyfrifol.

Cefndir

Mae deallusrwydd artiffisial (AI) yn datblygu'n gyflym ledled y byd, ac yn cynnig cyfleoedd sylweddol ond hefyd godi cwestiynau am ddiogelwch, preifatrwydd a rheoleiddio. Yn ddiweddar, mae nifer o arbenigwyr a llunwyr polisi rhyngwladol wedi galw am fwy o oruchwyliaeth er mwyn sicrhau bod AI yn cael ei ddatblygu a'i ddefnyddio mewn ffordd ddiogel, foesegol ac atebol.

Yn erbyn y cefndir hwn, cyhoeddodd Llywodraeth Cymru a Cosine eu bod yn cydweithio i gryfhau deallusrwydd artiffisial Cymraeg ym mis Medi 2026. Mae'r bartneriaeth yn seiliedig ar rannu arbenigedd a chyfeirio at adnoddau Cymraeg agored sydd eisoes ar gael yn gyhoeddus yn rhad ac am ddim. Does dim

ymrwymiad ariannol gan Lywodraeth Cymru o dan y trefniant hwn, a does dim trosglwyddo perchnogaeth unrhyw setiau data nac adnoddau iaith i Cosine.

Mae'r gwaith hwn yn adeiladu ar fuddsoddiad parhaus Llywodraeth Cymru mewn technoleg Gymraeg e.e. drwy Brifysgol Bangor, gan gynnwys partneriaeth flaenorol gydag OpenAI yn 2024.

Swyddog: [WG Official 1] ac [WG Official 2] 22/09/2026)

ATISN 27572 - Email - 20260813 - "FW: ATB: RE: Potential Cosine partnership announcement" - attachment

Cosine and Llywodraeth Cymru (Welsh Government) collaborate to support Welsh language representation in Lumen model training

The collaboration will bring Welsh language data into Cosine's Lumen training pipeline as part of a broader effort to build AI systems that better represent the UK.

[London/Cardiff], [Date] — Cosine today announced a collaboration with Llywodraeth Cymru (Welsh Government) to support the inclusion of Welsh language data in training for its Lumen Sovereign model.

The work will help ensure Lumen Sovereign reflects the UK more accurately in practice. That includes Welsh, one of the UK's living national languages and an important part of public, cultural, and institutional life in Wales.

As AI systems become part of everyday software, services, and digital infrastructure, language coverage matters. Models used across the UK should represent the languages and contexts that shape life here. This collaboration with the Welsh Government brings Welsh into that work in a direct and practical way and is part of a broader effort to build models that are more useful, better informed, and more representative of the country as a whole.

Alistair Pullen, CEO, Cosine

"Welsh is an essential part of the UK's linguistic landscape. This work with the Welsh Government helps us build Lumen on a stronger foundation, with better representation of the languages and public context that matter across the UK."

[Welsh Government spokesperson quote placeholder]

"The Welsh language is a vital part of Wales's identity and public life. Supporting its

inclusion in AI model development is an important step in ensuring new technologies better represent the people and languages of the UK.”

This collaboration sits within a wider effort to make Lumen more representative of the UK, including the languages, knowledge, and public context that general-purpose model development often overlooks.

Further details on the collaboration will be shared in due course.

About Cosine

Cosine is the British sovereign AI lab building frontier models for defence, national security, and highly regulated industries. Our Lumen model family is designed for software engineering and other high-value tool-using work in fully air-gapped, high-security environments.

Powered by our in-house ML research lab and backed by the UK Government’s Sovereign AI initiative, we are building the UK’s first fully sovereign frontier AI model. While we also post-train strong open-weight models on proprietary human-reasoning and domain-specific data, our flagship sovereign model is being developed with zero dependence on foreign infrastructure. This gives customers frontier capability with stronger control over training, deployment, and long-context performance.

Deployable entirely within a customer’s VPC or fully on-premise with zero data egress, Cosine gives governments, enterprises, and critical infrastructure providers sovereign control over the models, infrastructure, and data behind their AI systems.

ATISN 27572 - Email - 20260907 - "For input today - PN On Cosine partnership" - attachment

Welsh Government partners with Cosine to advance Welsh language AI

- New partnership with Cosine to help improve how artificial intelligence (AI) works in Welsh.
- Welsh Government to provide data, resources and expertise to support Welsh language AI.
- Improving technology to help more people to use Welsh in everyday life.

The Welsh Government has announced a new partnership with sovereign AI company Cosine to help improve how AI works in Welsh.

The Welsh Government will provide data, resources and expertise to help improve how Cosine's Lumen AI model understands and processes Welsh.

The partnership supports the Welsh Government's aim to ensure that Welsh thrives in the digital age, making it easier for people to use more of their Welsh in their everyday lives.

It also supports the government's work to ensure new technologies recognise, understand and support bilingual use, reflecting how multilingual communities all around the world use AI every day.

Chief Executive Officer of Cosine, Alistair Pullen, said: "Welsh is an essential part of the UK's linguistic landscape. This work with the Welsh Government helps us build Lumen on a stronger foundation, with better representation of the languages and public context that matter across the UK."

The Cabinet Minister for Education and the Welsh Language, Anna Brychan, said: "The Welsh language belongs to us all. It is part of who we are as a nation, and we want to create more opportunities for people to use it in every part of their lives.

"It's important that Welsh is fully represented in the technologies people use to access information, learn, work and communicate. This new partnership helps to strengthen Welsh-language capability in AI, creating more opportunities for us all to use our Welsh confidently every day and supporting the language's future in the digital age."

The Cabinet Minister for Enterprise, Connectivity and Energy, Adam Price, said: "AI has the potential to bring real benefits to our economy, public services and communities. By working with sovereign AI companies like Cosine, we can help build the skills, knowledge and capability Wales needs to make the most of these technologies, while supporting innovation that reflects our national priorities.

This will help strengthen Wales' long-term position in AI, and is especially important for the Welsh language, helping to ensure new technologies work for people in the language of their choice."

Ends.

Llywodraeth Cymru dewis Cosine fel partner i symud ymlaen â Deallusrwydd Artiffisial Cymraeg

- Mae Llywodraeth Cymru wedi cyhoeddi partneriaeth newydd â Cosine i helpu i wella sut mae deallusrwydd artiffisial yn gweithio yn Gymraeg.
- Bydd Llywodraeth Cymru yn darparu data, adnoddau ac arbenigedd i gefnogi deallusrwydd artiffisial Cymraeg.

- Nod hyn yw gwella technoleg i helpu mwy o bobl i ddefnyddio'r Gymraeg mewn bywyd bob dydd.

Mae Llywodraeth Cymru wedi cyhoeddi partneriaeth newydd gyda'r cwmni deallusrwydd artiffisial sofran Cosine i helpu i wella sut mae deallusrwydd artiffisial yn gweithio yn Gymraeg.

Bydd Llywodraeth Cymru yn darparu data, adnoddau ac arbenigedd i helpu i wella sut mae model Lumen AI Cosine yn deall ac yn prosesu'r Gymraeg.

Mae'r bartneriaeth yn cefnogi nod Llywodraeth Cymru i sicrhau bod y Gymraeg yn ffynnu yn yr oes ddigidol, gan ei gwneud hi'n haws i bobl ddefnyddio mwy o Gymraeg yn eu bywyd bob dydd.

Mae hefyd yn cefnogi gwaith y llywodraeth i sicrhau bod technolegau newydd yn cydnabod, deall a chefnogi dwyieithrwydd, gan adlewyrchu sut mae cymunedau amlieithog ledled y byd yn defnyddio deallusrwydd artiffisial bob dydd.

Dywedodd **Prif Swyddog Gweithredol Cosine, Alistair Pullen**: "Mae'r Gymraeg yn rhan hanfodol o gynfas ieithyddol y Deyrnas Unedig. Mae'r gwaith hwn gyda Llywodraeth Cymru yn ein helpu i adeiladu sylfaen gryfach i Lumen, a rhoi gwell cynrychiolaeth o'r ieithoedd a'r cyd-destun cyhoeddus sy'n bwysig ledled y Deyrnas Unedig."

Dywedodd y **Gweinidog Cabinet dros Addysg a'r Gymraeg, Anna Brychan**: "Mae'r Gymraeg yn perthyn inni i gyd. Mae'n rhan o'n hunaniaeth fel cenedl. Rydym am greu mwy o gyfleoedd i bobl ei ddefnyddio ym mhob agwedd ar fywyd.

"Mae'n bwysig bod y Gymraeg yn cael ei chynrychioli'n llawn yn y technolegau y mae pobl yn eu defnyddio i gael gafael ar wybodaeth, dysgu, gweithio a chyfathrebu. Mae'r bartneriaeth newydd hon yn helpu i gryfhau gallu'r Gymraeg mewn deallusrwydd artiffisial, gan greu mwy o gyfleoedd i bawb ddefnyddio'u Gymraeg yn hyderus bob dydd a chefnogi dyfodol yr iaith yn yr oes ddigidol sydd ohoni."

Dywedodd y **Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni, Adam Price**: "Mae potensial yma i ddeallusrwydd artiffisial ddod â manteision gwirioneddol i'n heconomi, gwasanaethau cyhoeddus a chymunedau. Drwy gydweithio â chwmnïau deallusrwydd artiffisial sofran fel Cosine, gallwn helpu i feithrin y sgiliau, yr wybodaeth a'r gallu sydd eu hangen ar Gymru i wneud y gorau o'r technolegau hyn, a hynny wrth gefnogi arloesedd sy'n adlewyrchu ein blaenoriaethau cenedlaethol.

Bydd hyn yn helpu i gryfhau sefyllfa hirdymor Cymru ym maes deallusrwydd artiffisial, ac mae hyn yn arbennig o bwysig i'r iaith Gymraeg, gan helpu i sicrhau bod technolegau newydd yn gweithio i bobl yn yr iaith y maent yn ei dewis."

Diwedd.

ATISN 27572 - Email - 20260910b - "RE: ATB/RE: ATB/RE: Potential Cosine partnership announcement" - attachment

Welsh Government partners with Cosine to advance Welsh language AI

- New partnership with Cosine to help improve how artificial intelligence (AI) works in Welsh.
- Welsh Government to provide data, resources and expertise to support Welsh language AI.
- Improving technology to help more people to use Welsh in everyday life.

The Welsh Government has announced a new partnership with sovereign AI company Cosine to help improve how AI works in Welsh.

The Welsh Government will provide data, resources and expertise to help improve how Cosine's Lumen AI model understands and processes Welsh.

The partnership supports the Welsh Government's aim to ensure that Welsh thrives in the digital age, making it easier for people to use more of their Welsh in their everyday lives.

It also supports the government's work to ensure new technologies recognise, understand and support bilingual use, reflecting how multilingual communities all around the world use AI every day.

Chief Executive Officer of Cosine, Alistair Pullen, said: "Welsh is an essential part of the UK's linguistic landscape. This work with the Welsh Government helps us build Lumen on a stronger foundation, with better representation of the languages and public context that matter across the UK."

The Cabinet Minister for Education and the Welsh Language, Anna Brychan, said: "The Welsh language belongs to us all. It is part of who we are as a nation, and we want to create more opportunities for people to use it in every part of their lives.

"It's important that Welsh is fully represented in the technologies people use to access information, learn, work and communicate. This new partnership helps to strengthen

Welsh-language capability in AI, creating more opportunities for us all to use our Welsh confidently every day and supporting the language's future in the digital age."

The Cabinet Minister for Enterprise, Connectivity and Energy, Adam Price, said: "AI has the potential to bring real benefits to our economy, public services and communities. By working with sovereign AI companies like Cosine, we can help build the skills, knowledge and capability Wales needs to make the most of these technologies, while supporting innovation that reflects our national priorities.

This will help strengthen Wales' long-term position in AI, and is especially important for the Welsh language, helping to ensure new technologies work for people in the language of their choice."

Ends.

Llywodraeth Cymru dewis Cosine fel partner i symud ymlaen â Deallusrwydd Artiffisial Cymraeg

- Mae Llywodraeth Cymru wedi cyhoeddi partneriaeth newydd â Cosine i helpu i wella sut mae deallusrwydd artiffisial yn gweithio yn Gymraeg.
- Bydd Llywodraeth Cymru yn darparu data, adnoddau ac arbenigedd i gefnogi deallusrwydd artiffisial Cymraeg.
- Nod hyn yw gwella technoleg i helpu mwy o bobl i ddefnyddio'r Gymraeg mewn bywyd bob dydd.

Mae Llywodraeth Cymru wedi cyhoeddi partneriaeth newydd gyda'r cwmni deallusrwydd artiffisial sofran Cosine i helpu i wella sut mae deallusrwydd artiffisial yn gweithio yn Gymraeg.

Bydd Llywodraeth Cymru yn darparu data, adnoddau ac arbenigedd i helpu i wella sut mae model Lumen AI Cosine yn deall ac yn prosesu'r Gymraeg.

Mae'r bartneriaeth yn cefnogi nod Llywodraeth Cymru i sicrhau bod y Gymraeg yn ffynnu yn yr oes ddigidol, gan ei gwneud hi'n haws i bobl ddefnyddio mwy o Gymraeg yn eu bywyd bob dydd.

Mae hefyd yn cefnogi gwaith y llywodraeth i sicrhau bod technolegau newydd yn cydnabod, deall a chefnogi dwyieithrwydd, gan adlewyrchu sut mae cymunedau amlieithog ledled y byd yn defnyddio deallusrwydd artiffisial bob dydd.

Dywedodd **Prif Swyddog Gweithredol Cosine, Alistair Pullen:** "Mae'r Gymraeg yn rhan hanfodol o gynfas ieithyddol y Deyrnas Unedig. Mae'r gwaith hwn gyda Llywodraeth Cymru yn ein helpu i adeiladu sylfaen gryfach i Lumen, a rhoi gwell cynrychiolaeth o'r ieithoedd a'r cyd-destun cyhoeddus sy'n bwysig ledled y Deyrnas Unedig."

Dywedodd y **Gweinidog Cabinet dros Addysg a'r Gymraeg, Anna Brychan**: "Mae'r Gymraeg yn perthyn inni i gyd. Mae'n rhan o'n hunaniaeth fel cenedl. Rydym am greu mwy o gyfleoedd i bobl ei ddefnyddio ym mhob agwedd ar fywyd.

"Mae'n bwysig bod y Gymraeg yn cael ei chynrychioli'n llawn yn y technolegau y mae pobl yn eu defnyddio i gael gafael ar wybodaeth, dysgu, gweithio a chyfathrebu. Mae'r bartneriaeth newydd hon yn helpu i gryfhau gallu'r Gymraeg mewn deallusrwydd artiffisial, gan greu mwy o gyfleoedd i bawb ddefnyddio'u Cymraeg yn hyderus bob dydd a chefnogi dyfodol yr iaith yn yr oes ddigidol sydd ohoni."

Dywedodd y **Gweinidog Cabinet dros Fenter, Cysylltedd ac Ynni, Adam Price**: "Mae potensial yma i ddeallusrwydd artiffisial ddod â manteision gwirioneddol i'n heconomi, gwasanaethau cyhoeddus a chymunedau. Drwy gydweithio â chwmnïau deallusrwydd artiffisial sofran fel Cosine, gallwn helpu i feithrin y sgiliau, yr wybodaeth a'r gallu sydd eu hangen ar Gymru i wneud y gorau o'r technolegau hyn, a hynny wrth gefnogi arloesedd sy'n adlewyrchu ein blaenoriaethau cenedlaethol.

Bydd hyn yn helpu i gryfhau sefyllfa hirdymor Cymru ym maes deallusrwydd artiffisial, ac mae hyn yn arbennig o bwysig i'r iaith Gymraeg, gan helpu i sicrhau bod technolegau newydd yn gweithio i bobl yn yr iaith y maent yn ei dewis."

Diwedd.

ATISN 27572 - Email - 20260923 - "CS1551 At Eich Sylw / For Your Attention" - attachment

CS1551: Cwyn safonau'r Gymraeg

Rydym wedi derbyn cwyn gan berson ar 17/09/2026. Mae'n ymwneud â chyhoeddiad Llywodraeth Cymru am ei chydweithrediad â Cosine. Mae'r achwynydd yn pryderu y bydd hyn yn "tanseilio budd y cyhoedd a chydaddoldeb ieithyddol hirdymor". Isod, dyfynnaf dri phwynt y mae'r achwynydd yn seilio ei gŵyn arnynt:

4. **Safonau Llunio Polisi (Safonau 88–93):** Methiant Llywodraeth Cymru i gynnal asesiad effaith iaith digonol cyn cytuno i drosglwyddo data

cyhoeddus a chorfforaethau iaith i gwmni masnachol preifat (Cosine). Mae perygl difrifol y bydd y penderfyniad hwn yn cael effaith andwyol ar gyfleoedd pobl i ddefnyddio'r Gymraeg, drwy greu monopolies masnachol a thalu ffioedd tanysgrifio/API am adnoddau a gafodd eu hariannu'n gyhoeddus.

Mae'r gŵyn yn seiliedig ar ddealltwriaeth fod Llywodraeth Cymru wedi cytuno i drosglwyddo data i Cosine. Mae'r term "partneriaeth" yn cyfeirio at ymgysylltiad cydweithredol â Cosine, yn seiliedig ar rannu arbenigedd a chyfeirio at adnoddau Cymraeg perthnasol sydd eisoes ar gael yn gyhoeddus ac sy'n *agored*. Roedd yr adnoddau hyn eisoes ar gael yn gyhoeddus ac maent yn parhau i fod ar gael i unrhyw un eu defnyddio a'u hailddefnyddio.

5. ***Egwyddor Trin y Gymraeg Ddim Llai Ffafirol na'r Saesneg:*** *Drwy beidio â sicrhau bod pwysau'r model (model weights) a'r cod yn agored (open-source / open-weights), mae Llywodraeth Cymru yn peryglu mynediad agored sefydliadau cyhoeddus, ysgolion, a llywodraeth leol at dechnoleg iaith Gymraeg sofran, tra bod adnoddau Saesneg enfawr eisoes ar gael am ddim yn fyd-eang.*

Drwy grantiau technoleg, mae Llywodraeth Cymru wedi, ac yn buddsoddi yn nhechnoleg ac AI Cymraeg, gan ariannu ystod o offer, data ac adnoddau sydd ar gael yn rhad ac am ddim i unrhyw un eu defnyddio a'u hailddefnyddio.

Gan fod Llywodraeth Cymru dim ond wedi cyfeirio Cosine at adnoddau agored, fel rhain, nid ydym yn penderfynu sut y gall Cosine, nac unrhyw sefydliad arall, aildefnyddio'r adnoddau agored hynny. Mae'r holl adnoddau'n parhau i fod ar gael yn gyhoeddus i unrhyw un eu defnyddio a'u hailddefnyddio.

6. ***Mynediad at Wasanaethau Cyhoeddus a Chyllid Trethdalwyr:*** *Mae'r cytundeb hwn yn amgáu comin diwylliannol cyhoeddus mewn cynnyrch caeedig, gan fygwth gallu cyrff cyhoeddus yng Nghymru i gydymffurfio â'u Safonau eu hunain yn y dyfodol heb orfod talu cwmni masnachol yn Llundain.*

Nid yw Llywodraeth Cymru wedi trosglwyddo perchnogaeth unrhyw setiau data neu adnoddau iaith i Cosine.

Mae'r adnoddau y cyfeiriwyd atynt fel rhan o'r bartneriaeth yn adnoddau agored sydd eisoes ar gael yn gyhoeddus ac sy'n parhau i fod ar gael i unrhyw unigolyn neu sefydliad o dan eu telerau a'u trwyddedau presennol.

Mae Llywodraeth Cymru yn glir bod y Gymraeg yn perthyn i ni i gyd ac na ddylai fod unrhyw rwystyr rhag defnyddio'r Gymraeg. Mae'r adnoddau technoleg Gymraeg a ariennir gan Lywodraeth Cymru ar gael yn rhad ac am ddim i unrhyw un eu defnyddio a'u haildefnyddio. Nid Llywodraeth Cymru sydd felly yn penderfynu sut y gall Cosine, neu unrhyw sefydliad arall, aildefnyddio'r adnoddau hynny.

Mae'r gŵyn yn ddilys sy'n golygu bod dyletswydd arnom i ystyried cynnal ymchwiliad dan adran 71 Mesur y Gymraeg.

Er mwyn gwneud yn siŵr bod gennym yr holl wybodaeth sydd ei angen, a dealltwriaeth lawn o'r sefyllfa cyn penderfynu, mae'n bwysig eich bod yn rhoi gymaint o wybodaeth a phosib i ni ar y cam hwn. Gall rhoi digon o wybodaeth i ni nawr olygu na fydd angen i ni gysylltu â chi eto am fwy o wybodaeth pe baem yn penderfynu ymchwilio.

A wnewch chi felly ymateb i'r e-bost hwn yn cadarnhau eich bod yn derbyn cyfrifoldeb am y mater dan sylw, gan roi gwybod i ni am:

- unrhyw amgylchiadau perthnasol sy'n esbonio beth sydd wedi digwydd

Daeth Cosine at Lywodraeth Cymru i geisio cyngor ac arbenigedd ym maes AI Cymraeg. Mewn ymateb, cyfeiriwyd Cosine at ystod o adnoddau agored sydd eisoes ar gael yn gyhoeddus a chytunwyd y byddai Llywodraeth Cymru yn parhau i rannu arbenigedd yn y maes hwnnw. Nid oes unrhyw ymrwymiad ariannol gan Lywodraeth Cymru.

Mae Llywodraeth Cymru yn parhau i gynnig cyngor ac arbenigedd ar dechnoleg ac AI Cymraeg i sefydliadau sydd â diddordeb yn y maes, gan gynnwys cyfeirio at adnoddau a data agored sydd ar gael yn gyhoeddus.

- unrhyw gamau rydych wedi eu cymryd yn barod i gywiro'r mater

Nid ydym yn ystyried bod angen cymryd camau cywiro gan fod y gŵyn yn seiliedig ar gamddealltwriaeth.

- unrhyw gamau y byddwch yn eu cymryd yn y dyfodol i geisio gwneud yn siŵr na fydd sefyllfa debyg yn codi eto.

Er mwyn osgoi camddealltwriaeth tebyg yn y dyfodol, byddwn yn fwy eglur mewn datganiadau a chyfathrebiadau cyhoeddus wrth gyfeirio at ddata neu adnoddau, gan nodi'n gliriach pan fo'r rhain yn adnoddau agored sydd eisoes ar gael yn gyhoeddus ac yn agored.

Anfonwch eich ymateb erbyn 02/10/2026 os gwelwch yn dda. Gallwch ofyn am fwy o amser os oes angen.

ATISN 27572 - Email - 20260721 - "RE: Welsh open data" – attachment

Welsh data sources

Retained material, approximate word volumes, recommended uses and future improvements

Source from the email	Retained rows / material	Approximate word volume	Most useful for	General feedback and future use
Cardiff University translation memory	1.39M Welsh CPT segments; 1.43M EN–CY pairs	24.2M Welsh words in CPT. Parallel: 24.8M Welsh + 23.2M English words	Large-scale Welsh CPT; translation mid-training; institutional terminology	The largest source by row count and one of the largest by words. Rights were clear under CC BY 4.0. The release is flattened, so it should remain segment-level rather than being joined into artificial documents.
Cwm Taf Morgannwg UHB translation memory	142k Welsh segments; 146k EN–CY pairs	2.52M Welsh words in CPT. Parallel: 2.56M Welsh + 2.31M English words	Health-domain CPT; medical translation; domain adaptation	A useful specialist source with strong health terminology. It needed PII and alignment checks but remained highly usable. Document IDs and original order would make future releases more versatile.
Senedd / Cofnod y Cynulliad translation-memory data	104k Welsh segments; 105k EN–CY pairs	22.2M Welsh words in CPT. Parallel: 22.2M Welsh + 21.1M English words	Parliamentary and policy Welsh; formal translation; future summarisation and QA	Much larger in words than its row count suggests because many records contain substantial passages. The supplied derivative is flattened; official Senedd Record XML would be preferable for long-context training because it preserves meetings, speakers and contribution order.
BydTermCymru translation memories	68.8k Welsh segments; 71.4k pairs; 1,409 long chunks from 78 reconstructed documents	1.36M Welsh words in CPT. Parallel: 1.39M Welsh + 1.34M English words. Long-form: 890k Welsh words	Government/legal CPT; translation; terminology; document modelling	One of the most versatile sources. OGL licensing was clear, translation quality was strong, and x-SourceFile metadata enabled genuine document reconstruction. The long-form figure overlaps partly with the segment data and should not be added to it as wholly new text.
Legislation.gov.uk Welsh/English data	52.6k Welsh segments; 64.6k pairs; 21.6k long chunks from 4,568 documents	1.05M Welsh words in segment CPT. Parallel: 1.27M Welsh + 1.24M English words. Long-form: 12.6M Welsh words	Structured long-form CPT; legal translation; legal retrieval, QA and summarisation	The strongest source for trustworthy document structure. XML preserves Acts, instruments, parts and sections. Its domain is narrow, so it should be mixed with broader Welsh.
Bangor Transcription Bank	46.1k cleaned transcript segments from about 60.8 hours	616k Welsh words	Spoken and colloquial CPT; future speech/language adaptation	Helpful for register diversity. Published order was randomised and session metadata was absent, so it cannot safely be reconstructed into conversations. Original order and speaker/session IDs would materially improve future releases.
Techiaith CC0 sentence and transcript resources	About 121k retained rows	About 1.22M Welsh words in total	Supplementary CPT; spoken variation; possible tagging or linguistic training	Clearly licensed and useful as supplementary data. About 1.01M words come from CoVoST-derived translationese, so that subset should receive a lower sampling weight than naturally authored Welsh.
Welsh Government website and publications	Broad crawl not completed	Not measured; likely several million to tens of millions of Welsh words	Modern policy and public-information CPT; summarisation; retrieval; later instruction data	Likely a strong source if page and publication boundaries are retained. Access was intermittently blocked, so we avoided producing a partial corpus. A first-party export or stable archive route would be valuable.
Welsh Government aligned website pairs	Roughly 490k candidate pairs, not bulk-admitted	Likely around 7–10M Welsh words, assuming normal government-document segment lengths; unverified	Translation mid-training	Attractive volume, but sample inspection showed noticeable alignment noise. A curated approved subset with document IDs and alignment-confidence scores would be much more useful.
Common Voice Welsh	About 118k validated distinct sentences; excluded	Approximately 1–1.5M Welsh words, estimated	Local speech work; ASR; speech-text evaluation	Useful text, but redistribution restrictions prevented inclusion in the published dataset. It also overlaps with CoVoST-derived material. Better handled as a separately governed local resource.
CorCenCC	About 11.2M published words; excluded	Approximately 11.2M Welsh words	Natural and spoken Welsh; research evaluation; non-commercial adaptation	Linguistically one of the best sources considered. CC BY-NC-SA terms do not fit unrestricted or commercial model training, but it could support a separate research-only experiment.
TermCymru CSV	Approximately 14 MB of terminology	Not meaningfully comparable to prose word counts; likely hundreds of thousands of lexical entries/terms	Terminology retrieval; constrained translation; glossary evaluation; post-training tools	Strong lexical resource, but not coherent prose. More valuable as a glossary, retrieval source, terminology benchmark or translation-time constraint set than as ordinary CPT.
Termiadur Addysg	One annual terminology archive; held	Not measured; likely tens to hundreds of thousands of terms rather than prose words	Education terminology and evaluation	Potentially useful for education-focused translation and post-training. The stated CC BY-SA licence did not identify a version clearly enough for this batch.

Source from the email	Retained rows / material	Approximate word volume	Most useful for	General feedback and future use
HeloBlod	No public bulk release	At least 1M translated words claimed	Modern business/SME translation; domain adaptation	Potentially valuable because it may contain more contemporary and commercial language than legislation. Use would require a direct data agreement, explicit licence and PII review.
Paidaruo	About 38 hours / 12.7k clips	Likely under 500k distinct Welsh words, with considerable prompt repetition	Speech modelling rather than text CPT	Repeated read prompts limit incremental text value. More useful for pronunciation, ASR or speech evaluation than continued language-model pretraining.
Welsh URL list, DigiGrid, Trawsgrifiwr, lemmatiser and embedding tools	No independent training corpus	No attributable corpus word total	Discovery, preprocessing and corpus mining	Useful pointers or tools, but they do not themselves confer rights to source text. Future source lists would be easier to operationalise with owner, licence, estimated size, format and redistribution status attached to every entry.

Other

- WQ100618(e), WQ100619(e), WQ100620(e), WQ100621(e), available at <https://record.assembly.wales/OrderPaper/WrittenQuestions/25-09-2026/>
- ATISN 27499, available at <https://www.gov.wales/freedom-of-information-responses>

