Adroddiad gan Lee Waters AS

3edd Gynhadledd Cymdeithas Seneddol y Gymanwlad:

Artificial Intelligence and Disinformation: 'Democracy in the age of deepfakes'

Singapore

18-20 Mehefin 2024.

Yn gryno

1. Mae deallusrwydd artiffisial (AI) yn declyn. Mae ganddo alluoedd malaen ac anfalaen gan ddibynnu ar y ffordd y mae'n cael ei ddylunio a'i ddefnyddio. Ni allwn ei atal, ond gallwn geisio ei lywio.
2. Mae soffistigedigrwydd AI yn newid yn gyflym ac yn rhagori ar dechnoleg reoleiddio neu ganfod. Ond mae ymwybyddiaeth isel o gyflymder y newid.
3. Potensial sylweddol i ddefnyddio AI i roi hwb i gynhyrchiant a chreu twf economaidd. Bydd swyddi'n cael eu disodli gan dechnoleg fwy effeithlon a bydd swyddi a meysydd gweithgarwch newydd hefyd yn cael eu creu.
4. Rôl weithredol i'r Llywodraeth reoli 'trawsnewid cyfiawn' mewn economi, sicrhau bod buddion yn cael eu lledaenu, bod pobl yn cael eu hailsgilio ac y defnyddir rheoleiddio i sicrhau tegwch a moeseg o ran sut y cymhwysir AI - OND mae technoleg yn rhagori ar allu'r Wladwriaeth i ymateb.
5. Mae Singapore yn mabwysiadu dull strategol glyfar o arwain y maes yn AI. Mae'n cymryd camau i fod ar flaen y gad ac mae llawer y gallwn ddysgu ohono a chydweithio ag ef.
6. Mae defnydd eang o ddelweddau ffug hynod real ar gyfer twyll (hawliadau yswiriant, dogfennau adnabod ffug), ac mewn pornograffi lle y mae defnyddio ffugiadau dwfn heb ganiatâd yn fath newydd o drais yn erbyn menywod.
7. Mae'n anodd iawn dweud a yw testun wedi'i gynhyrchu gan AI; dim ond siawns 50/50 o sylwi a yw llun yn real neu'n ffug, ac ni all hyd yn oed aelodau agos o'r teulu ddweud gwahaniaeth rhwng sain real neu ffugiad dwfn.
8. Mae AI ffugiad dwfn eisoes yn cael ei ddefnyddio i ddylanwadu ar etholiadau a siarad camwybodaeth mewn etholiadau.
9. Amcangyfrifir, o fewn tair blynedd, y bydd 90 y cant o'r holl gynnwys ar-lein wedi'i drin yn gyfan gwbl neu'n rhannol.
10. Mae'r UE yn ceisio rheoleiddio (fel y gwnaeth gyda GDPR) ond mae'r pum cwmni technoleg mawr sy'n dominyddu yn gwrthsefyll. Tsieina sy'n arwain ar wreiddio safonau AI yn y nwyddau y mae'n eu cynhyrchu, e.e. cerbydau awtonomaidd, i ffurfio'r dulliau a fabwysiedir yn y farchnad.

CLUSTFEINIWCH!

Yn yr etholiadau Indiaidd diweddar rhoddodd merch pennaeth milwrol y Teigrod Tamil araith, wedi'i ffrydio'n fyw ar YouTube, yn annog Tamiliaid ar draws y byd i ymgymryd â brwydr wleidyddol dros eu rhyddid.

Pa ots?

Bu farw 14 blynedd cyn ei darllediad ‘byw’ ac nid yw’n hysbys ei bod wedi dweud yr un o’r pethau hynny.

Defnyddiwyd negeseuon sain gan Joe Biden i annog cefnogwyr i beidio â phleidleisio yn rhagetholiad New Hampshire eleni. Nid ef ydoedd, ond recordiad hynod realistig yn defnyddio clôn sain o’r Arlywydd i atal pleidleiswyr rhag pleidleisio er budd ei wrthwynebwyr.

Dylanwadwyd ar etholiadau yn Slofacia gan sain ffug o un o’r prif ymgeiswyr yn siarad am godi cost cwrw!

Ac yn etholiad Maer Llundain eleni, clywodd cannoedd o filoedd o bobl sain o Sadiq Khan yn gwneud sylwadau llidiol cyn Diwrnod y Cadoediad a oedd bron ag achosi “anhrefn ddifrifol”. Roedd yn ffug.

Mae uwch-swyddogion diogelwch cenedlaethol yn yr Unol Daleithiau wedi bod yn paratoi ar gyfer “ffugiadau dwfn” i greu dryswch ymhlith pleidleiswyr mewn ffordd nas gwelwyd o’r blaen. Mae awdurdodau’r Unol Daleithiau’n ymwneud â gwaith cynllunio wrth gefn ar gyfer llywodraeth dramor o bosibl yn defnyddio AI i ymyrryd yn yr etholiad Arlywyddol.

Mae arbenigwyr yn amcangyfrif, o fewn tair blynedd, y bydd 90 y cant o’r holl gynnwys ar-lein wedi’i drin yn gyfan gwbl neu’n rhannol.

GWELD YW CREDU? IE?

Ym, nage.

Mae soffistigedigrwydd technoleg AI yn tyfu’n gyflymach gyflymach ac, ochr yn ochr â hynny, mae'r costau a’r rhwystrau i fynediad yn gostwng.

Mewn ychydig flynyddoedd yn unig mae deallusrwydd artiffisial wedi neidio o ‘AI cul’ sy’n trefnu cynnwys presennol, i AI ‘cynhyrchiol’ a all droi mewnbynnau testun yn ddelwedd, troi delwedd yn gân, neu droi fideo’n destun.

Mae creu ffugiadau dwfn AI (cyfuniad o ‘ddysgu dwfn’ a ‘ffug’) yn galluogi fideos ffug hynod realistig i gael eu gwneud sy'n gallu darlunio pobl yn gwneud pethau nad ydynt erioed wedi'u gwneud na’u dweud. Nid yw’r syniad mai ‘gweld yw credu’ yn wir mwyach.

Nid yw bellach yn bosibl dweud a yw testun wedi’i gynhyrchu gan ddeallusrwydd artiffisial ai peidio. Mae’r defnydd o Fodelau Iaith Mawr, fel ChatGBT, hefyd yn gallu

cynhyrchu traethodau myfyrwyr sydd bron yn anghanfyddadwy, gyda 94 y cant heb godi pryderon gyda marcwyr.

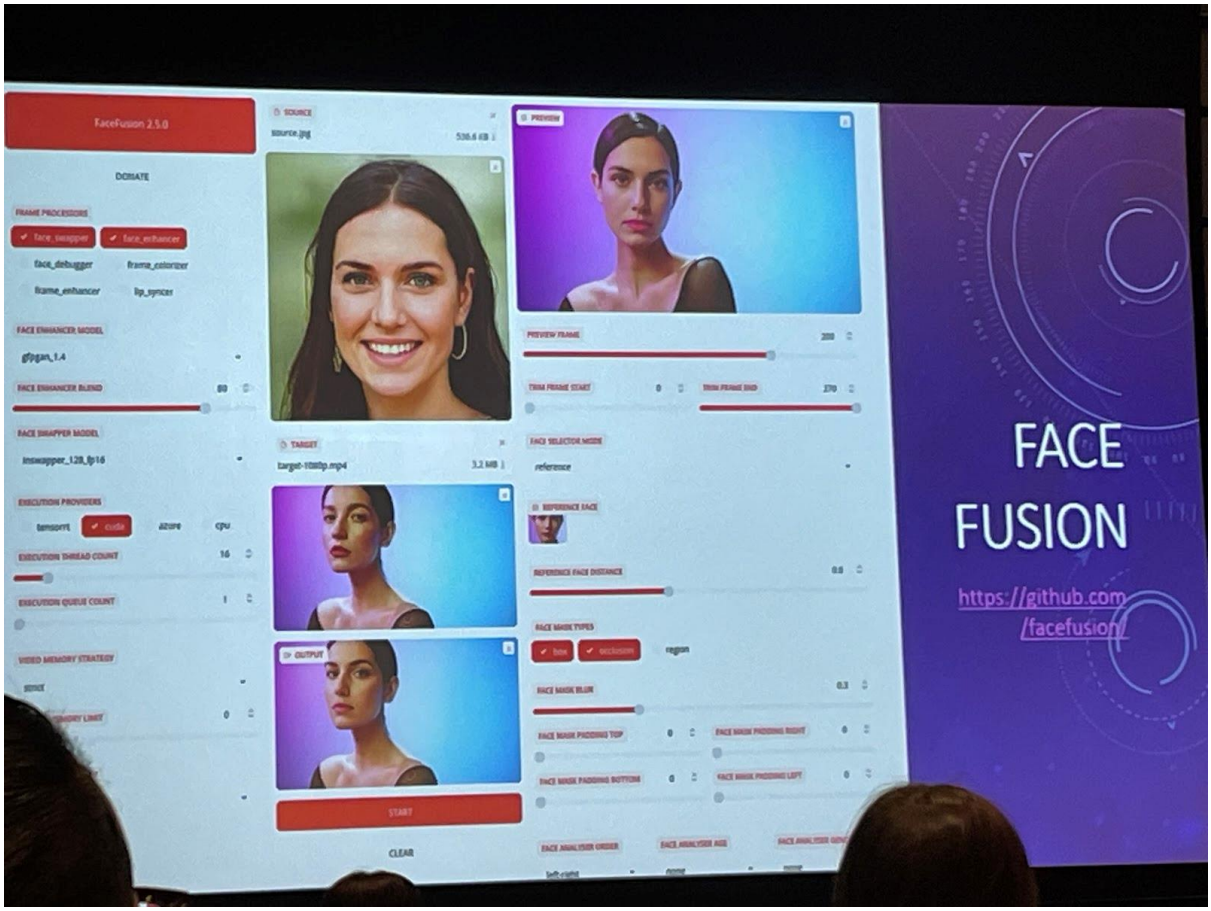
Dim ond hanner yr amser y gall rhywun ddweud y gwahaniaeth rhwng llun real neu ffug.

Ac nid yw hyd yn oed aelodau teulu agos yn gallu canfod ai chi sy'n gwneud galwad ffôn neu'n gadael neges sain

Nid peth newydd mewn etholiadau yw sïon ffug, camwybodaeth a thwyllwybodaeth. Ond mae'r cyfryngau cymdeithasol ac apiau negeseuon yn rhoi mecanweithiau danfon cyflym digynsail iddynt ar raddfa fawr.

Mae'n anodd olrhain ffugiadau dwfn, mae ganddynt gyrhaeddiad eang ac mae ymwybyddiaeth gyhoeddus isel o'r ddichell sy'n galluogi pobl i'w llyncu. Ac felly y mae cyflymder datblygiadau ym maes yr hyn a elwir yn 'AI Cynhyrchiol' nad yw'r gyfraith a rheoliadau'n gallu cadw i fyny. Ar ben hynny, mae ein synhwyrau a'n greddfau ein hunain yn ein methu hefyd.

Y cyfan y mae arnoch ei angen yw sampl o sain dair eiliad neu un ddelwedd ac mae VASA Microsoft yn gallu cynhyrchu symudiadau gwefusau wedi'u cydamseru'n gain â'r sain, yn ogystal â chipio sbectrwm mawr o naws wyneb a symudiadau pen naturiol sy'n cyfrannu i ganfyddiad dilysrwydd a bywiogrwydd.



‘ACTORION GWAEL’

Mae graddfa a soffistigedigrwydd y dechnoleg wedi cyrraedd y pwynt lle y gall lethu galluoedd ‘gwirio ffeithiau’.

Ni all technoleg bresennol ganfod ffugiadau dwfn sydd wedi’u creu gan dechnegau na chafodd ei hyfforddi arnynt, ac mae’n cael trafferth gwahaniaethu rhwng cynnwys real a ffug; mae'r dechnoleg ganfod sy’n dod i’r amlwg yn cynhyrchu nifer fawr o larymau ffug.

Mae ganddo hefyd gymwysiadau troseddol clir; gellir defnyddio delweddau ffugiad dwfn i dwyllo prosesau ‘cynefino’ fel ceisiadau pasbort, mesurau gwrth-dwyll banc ar-lein, a hawliadau yswiriant ffug.

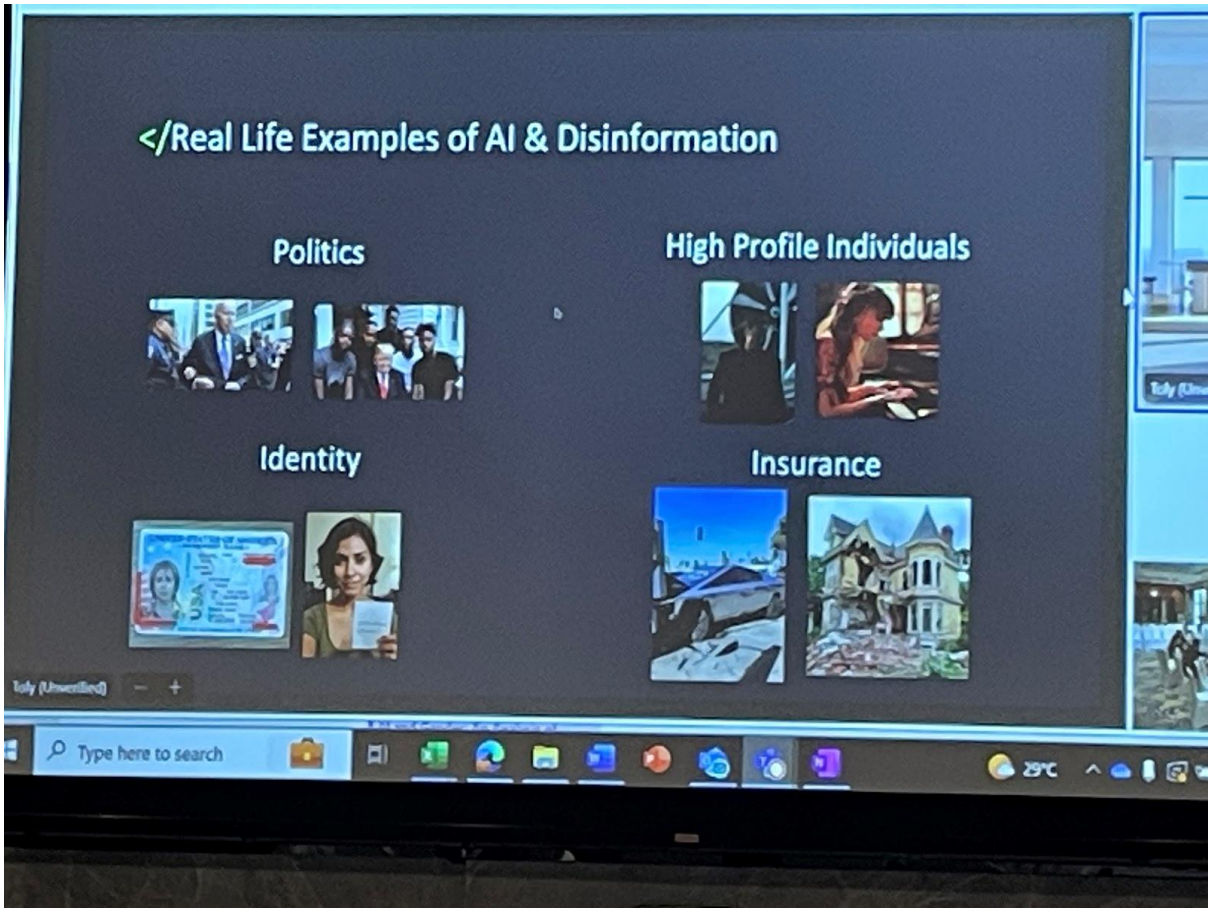
Gellir ei ddefnyddio hefyd ar alwadau Zoom/Teams i alluogi rhywun i esgus bod yn rhywun arall!

Bydd hyn yn arwain at bobl yn amau beth y maent yn ei weld a beth y maent yn gallu ymddiried ynddo. Gall hyn ‘wenwyno’r ffynnon’ ac arwain at 'ddifidend celwyddwr' fel y’i gelwir, pan ddefnyddir amheuaeth ynghylch ffugiadau dwfn i fwrw amheuaeth ar

dystiolaeth ddilys. Er enghraifft, ceisiodd dau ddiffynnydd a oedd ar brawf ar gyfer yr ymosodiad ar 6 Ionawr ar brifddinas yr Unol Daleithiau ddadlau yn erbyn fideo yn eu dangos yn y Capitol ar y sail y gallai fod wedi'i gynhyrchu gan AI.

Bydd hyn yn meithrin sinigiaeth ac yn tanseilio ymdeimlad cyfunol o realiti. Beth sy'n real? Beth sy'n ffug? Pwy y gellir ymddiried ynddo?

Mae arbenigwyr yn rhybuddio bod grym perswâd cynnwys hynod real - ond ffug, a'r ffaith y gellir ei bersonoli a'i ficrodargedu, yn golygu ei fod yn debygol o gael ei ddefnyddio i 'bregethu i'r rhai sydd wedi cael troëdigaeth' a dyfnhau'r duedd bresennol tuag at ragfarn cadarnhau. Gallai hyn atgyfnerthu'r ymdeimlad cynyddol o begynu a rhaniadau ideolegol.



TUEDDFRYD YN DYFNHAU

Mae elfen ryweddol benodol i ddatblygiad AI hefyd. Dim ond 12 y cant o ymchwilwyr AI a 6 y cant o ddatblygwyr meddalwedd proffesiynol ym maes AI sy'n fenywod. Mae'r diffyg amrywiaeth mewn timau datblygu AI yn cyfrannu i algorithmau rhagfarnllyd. Canfu astudiaeth yn dadansoddi 133 o systemau AI ar draws diwydiannau gwahanol fod tua 44 y cant ohonynt yn arddangos rhagfarn rhywedd, a bod 25 y cant yn dangos rhagfarn rhywedd a hil.

Mae dynion yn fwy tebygol o ddefnyddio AI yn eu bywydau proffesiynol neu bersonol, (mae 54 y cant o ddynion yn defnyddio AI a dim ond 35 y cant o fenywod sy'n ei ddefnyddio); ac efallai y bydd menywod yn fwy amharod i ddefnyddio teclynnau AI oherwydd pryderon am ymddiriedaeth, cywirdeb a llên-ladrad.

Canfu ymchwilwyr fod llawer o fenywod yn dewis cyfyngu ar eu gweledded ar-lein neu dynnu'n ôl yn gyfan gwbl i amddiffyn eu hunain, a all rwystro camu ymlaen mewn gyrfa, cyfleoedd rhwydweithio a thwf proffesiynol. Mae camdriniaeth o'r fath hefyd yn achosi i lawer o fenywod osgoi'r byd cyhoeddus, gan leihau lleisiau menywod mewn trafodaethau gwleidyddol ac arwain at benderfyniadau llai cynrychiadol a chynhwysol.

Mae algorithmau AI yn dysgu o batrymau yn y data, mae modelau AI yn tueddu i ddefnyddio'r mwyafrif fel y man cyfeirio er anfantais grwpiau lleiafrifol. Mae gwaith yn mynd rhagddo i ymchwilio i ffyrdd effeithiol o fynd i'r afael â mater 'tegwch algorithmig', ond mae academyddion yn rhybuddio y gall olygu cyfnewid a chyfaddawdu rhwng cywirdeb a thegwch.

Mae pryder cynyddol am ddefnyddio ffugiadau dwfn fel math o drais yn erbyn menywod. Amcangyfrifir bod 98 y cant o'r holl fideos ffugiad dwfn yn bornograffig ac mai menywod oedd testun 99 y cant ohonynt. Mae'r 'porn heb gydsyniad' hwn yn achosi cymaint o niwed â dosbarthu delweddau personol heb gydsyniad.

MAE ARNOM ANGEN RHEOLAU

Mae ymdrechion ar droed i reoleiddio'r defnydd o ddeallusrwydd artiffisial.

Mae rhai o'r cewri technoleg yn gwthio yn ôl ac mae'n well ganddynt hunanreoleiddio. Mae hyn yn hanfodol oherwydd mai dim ond 5 cwmni technoleg mawr sy'n dominyddu'r dirwedd - Apple, Meta (Facebook), Alphabet (Google), Amazon, Microsoft.

Mae'r UE wedi pasio darn ysgubol o reoleiddio AI i geisio dylanwadu ar safonau byd-eang, yn union fel y gwnaeth yn llwyddiannus gyda GDPR. Mae Tsieina wedi gwneud cynnydd sylweddol mewn dulliau mowldio. Mae arloesi Tsieineaidd ar archwilio a datgelu mewn perthynas ag AI, yn ogystal â'r safonau sydd wedi'u gwreiddio yn ei allforion technoleg a alluogir gan AI fel cerbydau awtonomaidd neu diwtoriaid digidol, eisoes yn arwyddocaol yn fyd-eang.

I helpu sefydliadau democrataidd i ymateb, mae Sefydliad Gwladwriaethau America a Chymdeithas Seneddol y Gymanwlad wedi datblygu Parliamentary Handbook on Disinformation, AI and Synthetic Media. Mae'r llawlyfr yn cynnwys strategaethau ar gyfer mynd i'r afael â thwyllwybodaeth ac arweiniad ar sut y gall seneddwyr weithio gyda chymdeithas sifil, y cyfryngau a chwmnïau technoleg, i ddatblygu fframweithiau rheoleiddio/deddfwriaethol i fynd i'r afael â heriau twyllwybodaeth, yn ogystal â sut y gallant gymryd camau i ddiogelu eu proffiliau ar-lein a'u sianeli cyfathrebu eu hunain.

NID YW'R CYFAN YN DDRWG

Mae Microsoft yn cydnabod y gellir camddefnyddio ei AI ffugiadau dwfn - FASA - ond mae'n dweud bod ganddo 'botensial cadarnhaol sylweddol' hefyd:

"The benefits – such as enhancing educational equity, improving accessibility for individuals with communication challenges, offering companionship or therapeutic support to those in need, among many others – underscore the importance of our research and other related explorations".

Yn yr etholiadau Indiaidd diweddar, yn ogystal â lledaenu gwybodaeth ffug helpodd botiaid AI hefyd i leihau rhwystrau ieithyddol drwy alluogi ymgeiswyr i gyrraedd mwy o bleidleiswyr sy'n siarad un o ieithoedd rhanbarthol niferus India.

Er enghraifft, defnyddiodd y Prif Weinidog Modi Bhashini, sef offeryn y llywodraeth wedi'i bweru gan AI, i sicrhau y gallai cynulleidfaoedd Tamil eu hiaith glywed ei araith, a gafodd ei thraddodi yn Hindi a'i chyfieithu i Damil mewn amser real. Mae ei areithiau hefyd wedi'u cyfieithu i Ganareg, Bengaleg, Telwgw, Odia, a Malaialam, ymhlith ieithoedd eraill, drwy ddefnyddio AI. Mae ap swyddogol y prif weinidog — NaMo — wedi lansio nodwedd a ddyluniwyd i hyrwyddo llwyddiannau polisi'r llywodraeth yn ehangach drwy sgwrsfotiaid wedi'u pweru gan AI.

Fel unrhyw offeryn, mae ganddo ddefnydd anfalaen a malaen.

YR ECONOMI, WRTH GWRS

Mae Singapore yn cymryd safbwynt strategol ar botensial economaidd cofleidio'r chwyldro AI. Lansiodd raglen ymchwil a datblygu genedlaethol yn 2017 a sefydliad hyd braich - AI Singapore- sef y dull cenedlaethol cyntaf o adeiladu galluoedd cenedlaethol dwfn ym maes AI. Maent yn canolbwyntio ar chwe maes strategol:

- Ymchwil - datblygu talent leol a chymhwyso AI i broblemau'r byd go iawn
- Technoleg - cefnogi prosiectau effaith uchel sy'n wynebu heriau cenedlaethol
- Arloesi - Sbarduno a chefnogi mabwysiadu AI yn eang ar draws y diwydiant
- Cynhyrchion - Adeiladu cymwysiadau ymarferol mewn meysydd datblygu allweddol
- Llywodraethu - Ymchwil mewn llywodraethu, moeseg ac atebolrwydd systemau AI

Mae'r dull hwn yn canolbwyntio'n helaeth ar waith ymchwil cymhwysol ochr yn ochr â'r sector preifat i heriau busnes y byd go iawn, a datblygu llif o dalent yn y wlad -

gan gynnwys cwrs Meistr dwys â thâl i dyfu'r gronfa dalent. Mae'r ymdrech ymchwil a datblygu yn canolbwyntio ar nifer fach o heriau strategol.

Rhai enghreifftiau ymarferol o brosiectau AI y maent wedi gweithio arnynt gyda BBaChau:

- Datblygu Model AI i gynorthwyo deintyddion drwy awtomeiddio proses siartio pelydr-X, gwella cywirdeb ac effeithlonrwydd. Mae'r arloesedd hwn yn galluogi deintyddion i ganolbwyntio ar ofal cleifion am fwy o amser.
- Cyd-gynhyrchu model i wella effeithlonrwydd cyflawni ar gyfer egin fusnes parseli yr un diwrnod. Gwellodd y model AI effeithlonrwydd llwybr 20 y cant a symleiddio gweithrediadau, gan wella ansawdd gwasanaethau a lleihau costau.
- Cyflymu dadansoddiad celloedd ar gyfer defnyddio microalgâu ym maes arloesedd bwyd o blanhigion. Mae cyfrif celloedd â llaw o dan ficrosgop i wirio parodrwydd cynaeafu yn cymryd llawer o amser (30-40 munud fesul sleid). Arweiniodd y cydweithrediad at ddatblygu teclyn AI i leihau'r amser dadansoddi 30 gwaith.
- Model Adnabod Llais ar gyfer galwadau brys sy'n helpu gweithredwyr drwy drawsgrifio mewn amser real, mewn sawl iaith i wneud y gorau o ddyraniad adnoddau a lleihau amseroedd ymateb.
- Cydweithio â phatholegwyr er mwyn gwella diagnosis o diwmor y fron, gan arwain at fodel CV pwrpasol 2 gam wedi'i hyfforddi ar ddelweddau meinwe. Gan gyflawni cywirdeb 87.5 y cant, mae'r offeryn yn cynorthwyo patholegwyr, yn torri costau ac ymyriadau llawfeddygol wrth leddfu gorbryder cleifion.

Mae disgrifiad llawn o'r dull y mae Singapore wedi'i ddilyn wedi'i ddilyn wedi'i gyhoeddi fel e-lyfr: AI-First Nation gan Laurence Liew.

Yn briodol defnyddiodd Gemini, fersiwn Google o Chat GBT, i ofyn cwestiynau iddo a oedd wedyn yn sail i'r llyfr - mae gan AI fywgraffydd, fel y dywedodd.

GADEWCH I NI SIARAD

Mae rhai sy'n credu y dylem roi'r gorau i ddatblygu AI ymhellach.

Mae Richard Heinberg, meddyliwr blaenllaw ar gynaliadwyedd, wedi dadlau bod rhaid rhoi terfyn ar ddeallusrwydd artiffisial nawr':

Pryd bynnag y cyflwynir technoleg newydd, yr arfer yw aros a gweld ei chanlyniadau cadarnhaol a negyddol cyn gweithredu rheoliadau. Ond os ydym yn aros nes bod AI wedi datblygu ymhellach, ni fyddwn wrth y llyw mwyach. Efallai y bydd yn amhosibl i ni adennill rheolaeth o'r dechnoleg yr ydym wedi'i chreu.

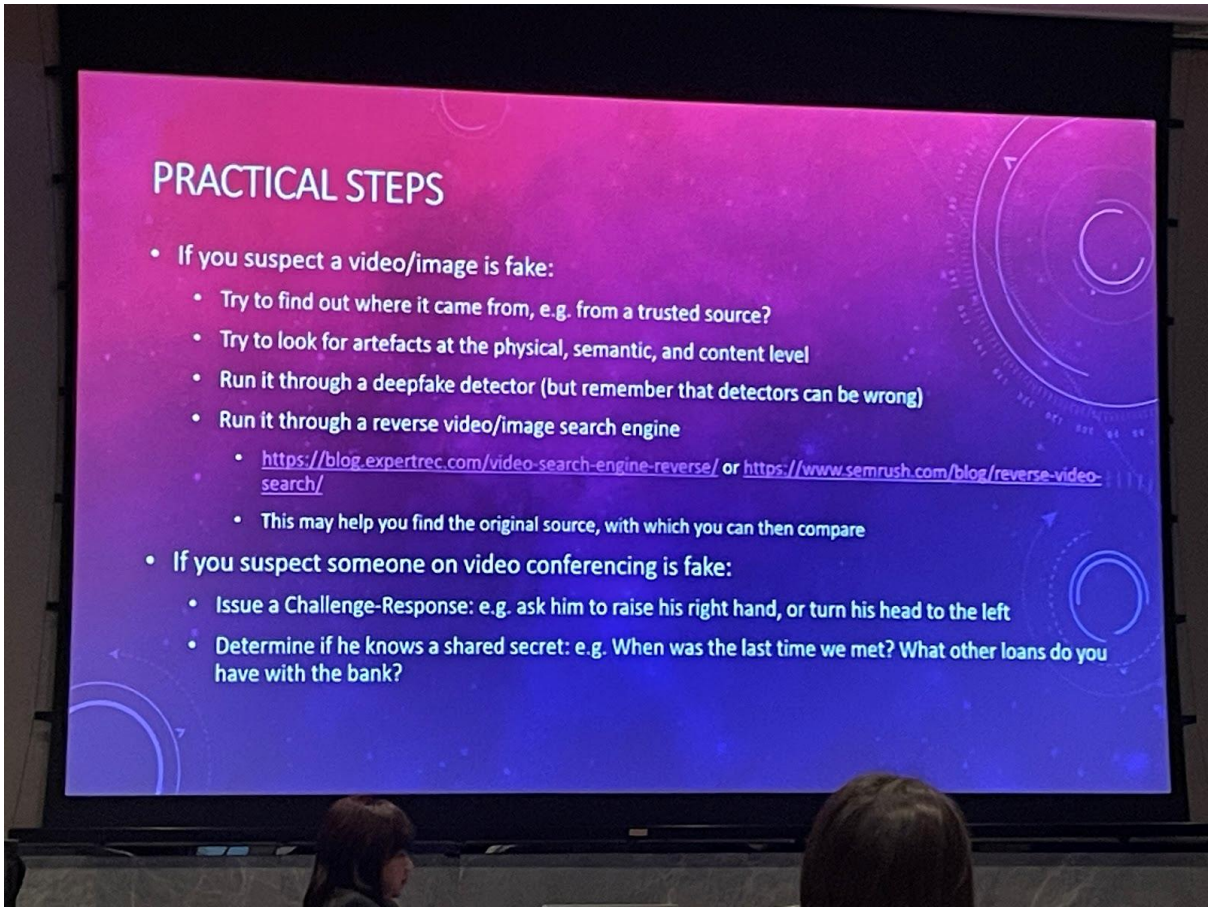
Rwy'n credu bod hynny'n ffansïol.

Fy marn i o'r dystiolaeth arbenigol gan arbenigwyr yn y gynhadledd yn Singapore, Rhaglen Ddatblygu'r CU, Awstralia a'r DU yw bod y llamau hyn mewn technoleg yn cynrychioli bygythiadau a chyfleoedd. Bydd yn anodd i dechnoleg a rheoliadau ei gadw dan reolaeth yn y tymor byr, ac un o'r pethau pwysicaf y gallwn ei wneud yw codi ymwybyddiaeth o'r hyn y mae AI yn gallu ei wneud, a sut y mae'n amlygu ei hun yn ein bywydau ac yn ein gwleidyddiaeth.

Gellir twyllo dulliau technegol megis dilysu cynnwys â dyfrnodau, neu ddatblygu technoleg a all wirio tarddiad cynnwys. Wrth i hynny barhau i ddatblygu, mae angen i bob un ohonom annog pleidleiswyr i ddechrau amau'r hyn y maent yn ei weld ac yn ei glywed.

Dyma'r consensws:

- Mae angen i sefydliadau democrataidd ddatblygu wrth i dechnoleg ddatblygu.
- Mae rhaid cael sgwrs gyhoeddus am sut y gall cymdeithasau ymateb i ddeallusrwydd artiffisial trawsnewidiol.
- Mae angen cydweithrediad byd-eang arnom i sicrhau bod AI yn ddiogel, ac i gyflawni buddion teg.



Cynhaliwyd 3edd Gynhadledd Cymdeithas Seneddol y Gymanwlad ar ddeallusrwydd artiffisial a thwyllwybodaeth: 'Democracy in the age of deepfakes' yn Singapore rhwng 18 ac 20 Mehefin 2024.

Yn bresennol yr oedd cynrychiolwyr o Borno (Nigeria); Ynysoedd Prydeinig y Wyryf; Cameroon; Ghana; Jamaica; Kenya; Malawi; Namibia; De Cymru Newydd (Awstralia); Nigeria; Penang (Malaysia); Queensland (Awstralia); Singapore; Sri Lanka; Y Deyrnas Unedig; Victoria (Awstralia); Cymru; Gorllewin Awstralia; Sansibar.

Lee Waters AS,
Cynrychiolydd Cangen y Senedd o Gymdeithas Seneddol y Gymanwlad.
Gorffennaf 2024