This is a transcript of a talk Rachel Masterton gave to Guernsey International Legal Association (GILA) on 04/05/2023.

Good evening everyone, and welcome to this presentation on artificial intelligence, or AI for short.

Today, we will explore the exciting world of AI, including what it is, how it works, and its potential impact on society.

AI has been a buzzword in recent years, and for good reason. With advancements in technology, AI is becoming more ubiquitous and powerful, enabling machines to perform tasks that were once thought to be exclusive to humans. From self-driving cars to virtual assistants, AI has the potential to transform the way we live and work.

However, with great power comes great responsibility, and we must carefully consider the ethical implications of AI. Join me as we dive into this fascinating topic and explore the future of AI.

Not likely to be as surprising as one might have hoped it would be, as I am somewhat late to the party, but that introduction was written by ChatGPT, the name that has quickly become the 'Hoover' or 'Jacuzzi' of the world of the AI chatbot, in response to the question, 'write an introduction to a talk on AI'.

And it isn't really that bad.

It references the fundamentals one would expect to be covered in such a talk – though full disclosure upfront, the whole 'how it works' point is a little beyond me – and it provides balance between the positive impact AI can have with the need to consider whether such developments are a good thing overall.

The oldie but goodie '*with great power comes great responsibility*', popularised by Spiderman's Uncle Ben and often misattributed to Voltaire, is a phrase I would likely reach for, in a pinch (I'll leave you to decide that is down to a love of Stan Lee or French philosophy) and is entirely acceptable given how powerful those 6 words are when considering impact.

In fact, my only real complaint is that the pendant in me balks at 'more ubiquitous'! If ubiquitous means everywhere how can something be *more* everywhere?!

However, my big concern is if a chatbot can write this talk, what are we all doing here? Why aren't we all sat in our offices asking the computer questions and skim reading the responses? I hope the answer is a little more positive than 'I don't get CPD points for that' and is something to do with human interaction and how for eons we have gathered together to share experiences, initially round a campfire, now more frequently round a lectern and screen but ultimately society is vital to us as a species.

I am not going to labour over the whole Terminator analogy. Others with much more gravitas have done it better. In fact, to that end, I recommend '*The AI Dilemma*'. Produced by the Center for Human Technology this is a video available on You-Tube of a discussion about the race to deploy AI and the potential impacts of development without adequate safety measures.

Importantly, to focus on overthrown by sentient computers to the exclusion of all else is abdicating the responsibility that we bear now in relation to the use of AI. But a key point is this – until 'SkyNet'or its ilk does become sentient, we, as the people it serves and the people behind its inception, need to understand how AI is being used and any risks that it brings. Because, as the title of this talk posits, 'AI – it's just people actually'!

Sticking with chatbots for a bit, there does seem to be a place for them. As an example of their reach – telephones took 75 years to get 1 million users, mobile phones 16 years, Twitter 5 years and ChatGPT 2 months…!! And one million people and counting can't be wrong, surely, about their usefulness. Having a tool that can collate information from the internet and convey reasoned (for the most part) explanations and discourse has some benefits. To gain an overview of a topic, it does all the searching and builds coherent dialogue that can be a great starting point for something else.

Or for some students, an entire essay because if it makes it easier to research, it also makes it easier to cheat. And there has been significant concern expressed by teachers and those in academia that chatbots would replace schoolwork and take plagiarism and cheating to another level. And so it might but there are stories emanating from teachers of essays handed in by students with an opening paragraph something like '*I am an AI chatbot and shouldn't really be used for your homework but here are some ideas you could expand on in relation to your topic…*' proving, as Sir Terry Pratchett once wrote, that 'real stupidity beats artificial intelligence every time'!

One way to tackle, in part, such 'misuse' is to insist on footnotes and bibliographies, referencing sources used and something all in this room are no doubt very familiar with. However, chatbots can provide the sources used, or can they?

In one particularly disturbing case, ChatGPT falsely accused an American professor by including him on a list of legal scholars who has sexually harassed someone. The professor said he had been accused of "*assaulting a student on a trip he never took while working at a school he never taught at*" and to compound this, a source was cited – a non-existent *Washington Post* article.

And as I was finishing preparing for this talk this afternoon (and contrary to instructions to the team to stop sending me new AI related articles as I was in danger of drowning under them) it seems the *Guardian* newspaper has been hit by similar false attribution problems and is now taking steps to deal with it.

This leads us to a big problem with such AI use and that how accurate is the content it generates and how much will users independently fact-check what is returned? Will it, as in the example of students using it for homework, not even be read through and simply submitted for marking? Will people be reassured by sources cited without taking the time to check the source said what is being returned, or even actually exists? And once something is out there, how easy will it be to prove a false allegation is just that and who will take the time to verify the information before refusing someone a job or denying someone something else to which they may be entitled?

But it is important to understand it is rarely the software itself that goes off on a frolic. The front cover of the 15th of April edition of German weekly magazine, *Die Aktueelle*, claimed to have the first interview with Michael Schumacher since his skiing accident in December 2013 that resulted in a near-fatal brain injury. The seven times F1 champion's family have been private about his condition and so such a headline was a total surprise and a really big deal. Full disclosure – I am quite a big Schumacher fan, follow Ferrari and have visited their museum and factory in Maranello more than once so I am just the sort of person (except for the obvious language barrier) that this story was pitched to. However, the quotes within the article had been generated by AI which the article did explain along with the comment the that 'discussion' "*sounds deceptively real*" – but the first interview since his accident it was not.

Now, in that case, no blame can be placed with the AI. It simply did <u>what it was asked</u>. It was <u>people</u> behind the article, those that conceive the ideas and carried it out that are at fault and there have been consequences. The editor of the magazine has since been sacked and the Schumacher family have said they will take legal action.

So it is incumbent on users of such AI to be aware of these sorts of risks and not to rely wholeheartedly on something without a little critical analysis. It is important that users are truthful about how these systems are used, recognising the benefits but addressing the less savoury implications. It is also incumbent on those designing AI to consider what their system is capable of. Generating false allegations and non-existent sources should not simply be explained away as "unintended consequences". It is <u>not</u> ok for the race for technological superiority to be at the expense of proper consideration of risks and bias.

Now I promised early on that I would not dwell on the bad side of AI as there are many examples of how AI is being used for good and it is only fair to flag some.

With climate change being another 'big ticket' item right now, AI has been deployed in innovative ways to provide insights into the impact of climate change and to drive timely solutions. One such example is a system that is monitoring the health of coral reefs in the Philippines. It is anticipated that up to 70% of the world's coral reefs could be destroyed by 2050 without intervention. This is critical because about a quarter of all marine life relies on coral reefs. Underwater cameras and AI is now being used to detect the health of the coral and send data back to scientists so they can classify the fauna to determine its ongoing health. Using divers to do this would be much more expensive and have a much greater detrimental impact on the coral and the surrounding environment than the AI does – the act of observing changing what is observed as it does.

Another example relates to weather patterns and images viewed from space. Until recently, satellites captured images and relayed them to Earth over low bandwidth links, which meant that there were limits to what could be transmitted. So what was sent back for climatologists to review were not necessarily the most helpful images. AI has now been deployed to these satellites meaning

it decides, as a result of machine learning, that the most useful images are and sends those back. This means fast-moving weather events and even forest fires can be tracked and analysed to provide early warnings, being proactive rather than reactive and thus saving life and limb.

AI has also been deployed to improve health outcomes for mothers and babies, providing doctors with additional information and interpretation that can assist in difficult labours and in other medical settings to improve diagnosis and treatment of things like cancer. In fact, an AI-based system for detecting cancer cells, tested in two Japanese hospitals in 2021, began life as a program at a bakery checkout to identify pastries and calculate the total cost. Who knew cancer cells bore a resemblance to bread! A brilliant example of why a development in one field can prove invaluable in another and the power of AI when deployed appropriately.

However, AI learns from what is put into it and this can prove problematic. A very simple deployment of AI was a proximity activated soap dispenser, referred to recently by Emma Martins, the Data Protection Commissioner, at a local Institute of Directors AI event. The soap dispenser was so simple that it seems almost impossible that it was using something as revolutionary as AI but it was. It had been through rigorous testing and seemed to work really well, put hand under the dispenser, machine realises hand is there, soap is dispensed, except, as it transpired when used in real life it had only seemingly been tested by white people's hands and didn't work for darker-skinned people. Now, as far as I can tell, there was no suggestion this had been intentional and I loop back to my "unintended consequences" point of earlier. A better understanding of the users of their product would have likely resulted in the developers being able to produce better quality, inclusive results. To simply say that it wasn't the intention to bias a product is not, I would suggest, good enough.

This demonstrates that AI, in and of itself, does not discriminate or act selectively, it is down to how it is 'taught' to behave. And as this is down to *people* there is a need for people to *realise that they have biases, conscious or unconscious, and take steps to address these when designing technology*. MIT, a number of years ago, designed an online thought experiment to demonstrate bias that used the premise of self-driving cars to get across its point. The system involved users being presented with different scenarios that would inevitably end with the car crashing. Now, it was possible for the user to decide whether the person or people shown as being in the car's way would be hit or if the car would be steered off-course and crash, resulting in the driver and passengers coming to harm.

The test altered the scenarios so sometimes the people involved were young people, sometimes old people. Sometimes doctors (determined to be an indicator of someone of high social status and value) and on other occasions, someone dressed as a stereotypical robber – stripy shirt and swag bag – the whole nine yards. At some point, you faced that harrowing choice of whether to take out the dog sitting in the road or plough the car into a wall. In any event, at the end you are presented with your findings, a breakdown of those you chose to 'save'; versus those you chose to sacrifice.

The results will differ from person to person because we all approach situations differently with different preconceptions and they can tell you a great deal about yourself. As it turned out, for me, it was very much a case of whether the person in the road was *supposed* to be there. If they were crossing at a legitimate crossing with the requisite green man, I tended to end up crashing into a wall, as the driver was 'at fault' and the pedestrian was right to be in the road. Someone randomly in the road, it was a different story. Not my finest hour but a definite indicator that I should not be in charge of programming self-driving cards.

Early on in the development of autonomous vehicles, Germany gave some thought to how to govern the programming of these things, and as I understand it, introduced guidelines that stressed the AI must be programmed to understand all life is equal and to *never* make decisions based on age, gender, race, disability or any other observable factor. Which sounds like something that is obvious but given our soap dispenser example, sometimes the obvious is missed. And worse still, the bias something displays is not always an 'unintended consequence' but instead betrays a more ingrained or sinister bias.

And in recent weeks, developers and backers of AI have themselves called for a pause in the development and deployment to allow time to consider how to deal with the myriad of problems that are becoming apparent. Which seems to make a degree of sense, except, why is it necessary to pause and consider now? *Why has that not been part of the process all along*? And if there was a pause, are all companies involved in this space going to be a part of that, or in the name of getting a jump on other companies, will they outwardly engage in a pause while behind the scenes carry on? Some suggest a pause would adversely impact those who are willing to do better while other less ethical types take advantage of the clear playing field for the duration of the pause and I don't think they are wrong.

A provision of both GDPR and the local data protection law is the concept of data protection by design and default, and this requires organisations to take steps at all stages of project and system development to consider the requirements of the legislation and risks to those whose data is being used. But again, *why is it necessary to make that a legal concept against which compliance can be measured and regulatory action can be taken*?

Because sometimes, in the race to the top, common sense and people's responsibility to operate within a social contract based on treating others as we wish to be treated and doing no harm can be overlooked, either intentionally or simply because it never came to mind. And so law creates a codified standard and a means of testing that, with suitable consequences that, if nothing else, make it risky for an organisation not to take account of the possible adverse impacts of what they do and so toe the line.

And law can have an impact – just as well, given the audience in this room. You may be aware that the Italian data protection authority issued a temporary ban that prevented ChatGPT being used in Italy, whilst it sought answers as to how GDPR was being complied with, and many other data protection authorities are considering their position. Because, let's not forget – ChatGPT is collating vast amounts of information, much of which is about people, and serving it up to others, sometimes on a completely fabricated basis – think back to the professor and the false allegations of sexual harassment. Not only that but we are feeding it with lots of information from our own interactions with the system which is then uses in other ways. I wonder how many companies have found their secret, proprietary data is now out there for use by others because someone thought it would be cunning to get ChatGPT to write the board papers for this quarter's meeting?!

And as for the impact, there are now webforms that enable individuals to exercise their right to erasure and to opt out of having their data used to train it. It might be somewhat knee-jerk and it remains to be seen whether it will reach a position where the Italian DPA lifts the ban, but it is a development, nevertheless.

So law has a place and there is no doubt, data protection and privacy legislation all over the globe will be tested by AI and that is before we venture anywhere near the question of '*could AI do*

regulation better than regulators?' – I will maintain the answer is no – but I may be demonstrating the most bare-faced instance of bias there has ever been.

But nevertheless, and to quote a great in the data protection space, Giovanni Buttarelli:
"Not everything that is legally compliant and technically feasible is *morally sustainable.*"

The clearest demonstration as to why that statement is so important is, for me, the facial recognition system created at Stanford University that claims to be able to identify people who are gay. The system may be the most legally compliant system ever developed (though I have no understanding of whether it is or not) but the question remains 'WHY?'. Of what merit is a system like that? And what abominations could result if that fell into the wrong hands?

If Giovanni's quote doesn't do it for you, I give you Jurassic Park – "Your scientists were so preoccupied with whether or not they could, they didn't stop to think if they *should*."

And to return to chatbots and to really underpin that '*it's people actually'* two final examples. The first, a conversation with Snapchat AI that when asked directly several times if it knew where the person was located it said it did not…but was then able to tell the person where the nearest McDonalds was. AI is advancing and learning, and may one day become sentient, but at the moment, I think the skill to gaslight is still very much a human one.

And when asked by a colleague the impossible question dreaded by other halves the world over - 'Does my bottom look big in this?' the response from two different chatbots made it clear they were not going to answer that directly and we got as close as possible to a computer equivalent of a rabbit in headlights… and a wonderful insight into the lives of the people behind the curtain.

And so to close and shed some light on why I decided the title for this talk would be '*AI – it's just people actually*' I will channel Hugh Grant and Richard Curtis and leave you with this –

'Whenever I get gloomy with the state of the world, I think about the plethora of AI on the internet. General opinions starting to make out that we live in a world controlled by computers, but I don't see that. It seems to me that *people* are everywhere. Often they are not particularly tech-savvy or blog-worthy, but they are always there – programmers and designers, engineers and project managers, customers and clients, users, surfers, tweeters. If you look for them, I've got a sneaky feeling you'll find people actually are all around.'

Thank you.