CHAT MAY BE A COPYCAT BUT WHAT ABOUT COPYRIGHT.

This article mainly focuses on the legal aspect of private property and how the creations of CHAT are challenging and redefining the concept of private property rights. Its lesser task is to debunk some of the myths of deep learning based on neural nets and the overblown claims it heralds unlimited learning.

The issue of copyright looms large with regard to CHAT both on the input side and the output side. Lawyers and governments are rushing behind these programs trying to catch up to the implications it poses to personal information and private property. The like of this legal turmoil has not been seen since the mid-19th Century when company law had to emerge to regulate the conduct of newly emerged joint stock corporations. On the input side CHAT harvests large swathes of information. Some of this is free information, some of this is protected information and some of it is copyrighted information. On the output side, the question is posed as to who owns the creations of CHAT. So let us begin.

The industrial revolution introducing large scale production requiring joint investment and ownership threw up many legal challenges on how best to operate and encourage these units of production. The result was the first of a series of company laws beginning with the *Joint Stock Companies Act of 1844* in Britain. In time the result was the emergence of corporations or companies with their own independent legal status. "*A corporation is a legal entity that is separate and distinct from its owners. Under the law, corporations possess many of the same rights and responsibilities as individuals. They can enter contracts, loan and borrow money, sue and be sued, hire employees, own assets, and pay taxes." (*Investopedia) The key to the success of the corporation however was 'limited liability' which anonymized the corporation and freed investors from being directly sued for the behaviour of the corporation. While there are exceptions for malicious and reckless behaviour, it does generally limit the liability of the owners to the extent of their investment, but not beyond it. On the other hand they stand to lose their entire investment.

But the important point is that limited liability allows investors to quantify their risk in advance, and risk and reward as we know is the binary code of capitalism. While equity investors are protected, customers and debtors are not protected on the downside as the Sackler family episode over their mis-selling of oxycodone showed when they stripped their equity out of *Purdue Corp* after it became clear the company was facing insurmountable legal challenges. In short, private property is surrounded by a maze of legality intended to regulate the gains and losses arising from the various forms of assets and debts so that naked competition does not tear the system apart. Such a crude mode of production complicated by competing interests patrolled by an army of lawyers remains wasteful of the labour time of society.

Copyright law emerges earlier in 1710 with the Statute of Anne in England. It reaches international maturity in 1886. "*In 1886, however, the Berne Convention was introduced to provide mutual recognition of copyright between nation states, and to promote the development of international standards for copyright protection. The Berne Convention does away with the need to register works separately in each individual country, and has been adopted by almost all the nations of the world (over 140 of the approximately 190 nation states of the world)."* Copyright law has become increasingly important in the information age.

In the end it boils down to who owns what and who can benefit from possessing it. In the words of Marx, property rights codify possession and regulates its claims on surplus value. It is the vagueness of these rights with regard to artificial intelligence today which constitutes the predicament facing lawyers and

legislators as they grapple with deep learning and the complexity of these new products. Deep learning may appear to create new knowledge but in fact it assembles and fashions old knowledge from the information it gathers on the internet. This old knowledge can be quite creative when re-ordered according to the mathematical rules that govern art itself. Thus CHAT can produce unique prose and music. Let us take music. While the combination of components may be unique, some of the components making up that piece of music could have preceded CHAT and they can be subject to copyright. If this is so and they are identified and dated by their originators, this can be construed as copyright infringement and thus be subject to penalties.

But who is to be sued. I have looked at a number of legal opinions. It seems that Microsoft and OpenAI cede their rights for the use of their CHAT products to the end user. And they do so on condition that the end user checks the output for accuracy, bias, and confidentiality. Here is a clause in the fine print of OPEN AI end user contract. *"The author generated this text in part with GPT-3, OpenAI's large-scale language-generation model. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication."* So the risk or damages arising from the output falls on the user and not the creators of the program.

Here is more detail taken from the most comprehensive opinion I have found outlining the legal risks when using CHAT. *IP Lawyer vs. ChatGPT: Top 10 Legal Issues of Using Generative AI at Work*.

It sets out the top ten risks which I will just headline.

1. Unauthorized use of Copyrighted Data to Train Generative AI Models.
2. Does the Output Violate Copyright Laws?
3. Confidentiality of Input of Prompts.
4. Data Ownership of the Prompt and the Output.
5. Authorship. (more on this later)
6. Seeking Copyright Protection on the Generative AI Content.
7. Bias in Outputs.
8. Factual Inaccuracies in Outputs
9. Algorithmic Disgorgement (penalize companies that have illegitimately collected data).
10. Indemnification (in case CHAT produces a copyrighted likeness the user indemnifies Open AI.)

Given this mine field composed of confidentiality and copyright it is no wonder that OpenAI cedes ownership of the results generated by CHAT to the end user. This from another legal firm Morgan Lewis: "*Content Ownership: 'Content' is defined in OpenAI's Terms of Use (Terms) for ChatGPT as anything input to or output generated by the service. With respect to ownership, the Terms state: "[a]s between the parties and to the extent permitted by applicable law, you own all Input, and subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output." OpenAI retains limited rights to use content for the purpose of providing and maintaining the service, complying with applicable law, and to enforce OpenAI's policies."* So the end user owns both the input and output.

The issue of deep learning and its output is most acute in the medical world where a mistaken diagnosis can be immediate, catastrophic, and irreversible. This article by Medscape, a Doctor's portal, outlines the dangers and risks when there is an over-reliance on algorithms to make a differential diagnosis. Legally, there can be no defense against a malpractice suit if it is shown the doctor did not make an independent assessment thereby checking the prompt from their computer. The article also discusses the issues of

patient confidentiality and latent biases in programs especially as they relate to the treatment of minorities. It also reveals how extensive the use of deep learning in the medical world already is.

Waiving their rights by the creators of these systems makes no commercial sense. Surely the creators of CHAT, who have invested billions of dollars and millions of hours in its production, would like to share in its spoils. For example OpenAI is expected to lose $540 million this year and is looking for ways to monetize its Intellectual Property. The fact that Open AI demurs in this case, shows it considers that potential losses could outweigh any gains, and the losses here are assumed to arise from future litigations over inputs and outputs. These conditions by OpenAI are more reminiscent of a fence dealing in stolen goods than a reputable software company. Never has this traditional maxim been more apposite: "*Buyer Beware".*

But there is more to come. In the first opinion cited above, in the USA at least, it is impossible to copyright the output of CHAT. "*For example, in the United States, copyright laws do not protect works created solely by a computer but works in which an individual can demonstrate substantial human involvement may qualify for copyright protection."* So in the USA a company cannot copyright and therefore monopolise an image for commercial gain or to prevent other users passing it off, but they can still be held accountable if that image or work contains previously copyrighted components. So no gain but a potential loss. (It is worth pointing out that in the UK computer output can be copyrighted whereas in the EU it is more nuanced.)

Finally, because CHAT continues to hoover up information including the output of users, companies generating material for their own use based on their own business experience and data may find these resources commandeered by CHAT and made available to their competitors. Thus businesses need to be cautious about inputting information or find ways to seal off their trade secrets from CHAT when using it, which of course makes it less effective as the key to the commercial success of CHAT is to tailor it to the company's own requirements. Sealing off information is not that simple because OpenAI, Microsoft and Google keep the physiognomy of their deep learning models a tightly held secret.

All in all, it will likely be the courts which will first refine the relationships between the creators of CHAT and end users until such time that statute law catches up. In the meantime end users could mount the defense that they cannot be held liable for information they were not party to copying, or whose source was kept hidden from them by the makers of CHAT or its equivalent. It will be messy and take time, but capitalism expects nothing less.

And it is not only the courts.

It will be workers as well. The attack on jobs by generative models is currently focused in the creative sphere, particularly on film making and scripting. Generative deep learning models can already prepare scripts and if not fully prepare them, then help complete them more quickly. This has led to a strike by the Writers Guild of America, now entering its second month, because of the potential cut in jobs and hours paid. According to a Financial Times article, actors are now poised to join the writers because of the intended use of deepfake replicas 'acting' on their behalf in films. Should they strike then Hollywood would ground to a halt showing that deep learning by workers as to the consequences of this innovation is itself a factor in the introduction of generative 'AI'.

The main concern of actors is the issue of informed consent before the studios can use their digital doubles to impersonate them in films. Clearly this is a monetary consideration as the actors would then benefit

jointly from the use of their digital doubles. As always, it is the issue of copyright, who owns the image and the right to use it. (Expect lots more stunts in future films, aka digital stuntmen, and women.) By using these digital doubles, studios which could only produce 2 films per year can now produce 6, because, while humans can do expressive slow acting, computers do compressed vivid fast acting. Thus while film productivity may improve it is bound to dumb down films. Dare I coin the term *fake acting*.

It is likely that actors and writers will not be the only workers thrown into the fray of class struggle. Deep learning as I have never tired of pointing out, targets office workers more than industrial workers. Though office workers have less of a tradition of militancy compared to shop floor industrial workers connected by the rhythm of production, the sheer scale of the attacks will force them to act. The Union Movement should prioritize and task themselves with recruiting and organizing these sections of workers.

The destabilizing political consequences of culling large swathes of office workers has not gone unnoticed by the silicon valley billionaires, hence the renewed interest in offering a *Universal Basic Income* to damp down unrest and disaffection.

The issue of productivity.

In the end capitalism only employs a new technology if it economizes on labour time sufficient to boost the rate of profit. The benefit to capitalists from this software is that for a small outlay, large improvements in productivity can be achieved. But as Marxists we should distinguish between productivity and productivity. Here we are dealing with a special case - economizing on the labour time of unproductive workers by reducing the expenditure of labour in this sphere.

If we take productivity at the level of the corporation to mean the following - output divided by the number of employed workers - productivity can increase even if output remains unchanged should fewer workers be employed. Fewer workers include thinning out office jobs, jobs belonging to the Loss side of the Profit and Loss Account. This means there will be fewer deductions from the gross profit created by productive workers which is transferred to the Profit and Loss Account. The result of this reduction in expenses will be a boost to the Net Profit transferred to the Balance Sheet (the capitalists' ledger recording their wealth and additions to their wealth). In short it will appear as though there is an increase in productivity because of the increase in net profit. But this is not an increase in the mass of profits but an increase in the residual profit belonging to the owners of the corporation courtesy of the reduction in expenses.

Now let us turn to the productive workers, the workers who actually produce the commodities that form the revenue of the corporation and with it the fountain of gross profit. Here there will be efficiencies especially when deep learning is attached to the Internet of Things (IoT), that is to say the real time monitoring of equipment via sensory inputs with the results being continuously analyzed. This will lead to improved designs, less breakdowns and less downtimes as maintenance can be fine tuned. However this is not new. Jet Engine manufacturers such as Rolls Royce and General Electric have been monitoring the performance of their engines in flight via satellite communications to ensure these engines are working at their optimum. What is different is that the advances in 5G and deep learning make possible the economic monitoring of much cheaper pieces of equipment. Hence making machinery 'clever' has been dubbed the Fourth Industrial Revolution.

One of the intriguing issues is whether or not deep learning can spit out unforeseen new products. This is highly unlikely. Deep learning cannot see that which does not yet exist. It is best suited to making existing

products better and with more features. Should designers solely rely on deep learning it can have a stupefying effect. On the other hand a new design can be introduced more quickly by using deep learning to analyze its features and functions. One concern remains, will deep learning uncover all the faults in a new product, or will overconfidence in this system and the rush to market result in a notable accident? Probably.

How quickly can capitalism expect this productivity boost? According to the Conference Board (TED) it will take up to 10 years. As the Table below shows it will barely have an effect this year. This means it will not yet be in a position to act as a counter-vailing factor to the undoubted increase in cost prices resulting from reshoring the international supply chain and the subsequent increase in costs due to the loss of least cost options, standardization, and economies of scale. Instead its impact will likely only emerge after the recession, not before it, because recessions have the tendency to accelerate the adoption of new technologies in order to restore the depressed rate of profit.

	2011-2019	2020	2021			2022			2023 (Forecast)		
	GDP per Hour Worked (avg)	GDP per Hour Worked	GDP	Total Hours Worked	GDP per Hour Worked	GDP	Total Hours Worked	GDP per Hour Worked	GDP*	Total Hours Worked	GDP per Hour Worked
US	+0.6%	+4.0%	+5.9%	+4.5%	+1.4%	+2.1%	+3.2%	-1.1%	+0.7%	+1.4%	-0.7%
Euro Area	+0.7%	+1.5%	+5.1%	+5.4%	-0.3%	+3.3%	+3.4%	-0.2%	+0.5%	+0.8%	-0.2%
Germany	+1.0%	+1.0%	+2.6%	+1.7%	+0.9%	+1.8%	+1.4%	+0.4%	+0.1%	+0.4%	-0.3%
France	+0.8%	+0.5%	+6.8%	+8.4%	-1.5%	+2.6%	+4.3%	-1.6%	+0.5%	+0.6%	-0.1%
UK	+0.5%	-0.6%	+7.6%	+6.6%	+0.9%	+4.1%	+3.6%	+0.4%	-0.2%	-0.1%	-0.1%
Japan	+0.9%	-1.4%	+2.1%	+0.2%	+2.0%	+1.0%	+0.3%	+0.8%	+0.7%	-1.6%	+2.3%
All Mature Economies	+1.0%	+2.0%	+5.4%	+4.0%	+1.4%	+2.7%	+2.8%	-0.1%	+0.7%	+0.4%	+0.4%
China	+7.6%	+3.5%	+8.4%	+0.3%	+8.1%	+3.0%	-3.4%	+6.7%	+5.1%	+0.3%	+4.8%
India	+6.3%	+6.2%	+9.6%	+11.3%	-1.5%	+6.9%	+8.3%	-1.3%	+5.0%	+1.6%	+3.3%
Latin America	+0.4%	+9.4%	+6.8%	+14.7%	-6.9%	+3.7%	+6.5%	-2.7%	+0.9%	+0.2%	+0.6%
All Emerging & Developing	+3.9%	+5.0%	+6.9%	+5.2%	+1.6%	+3.7%	+3.3%	+0.4%	+3.5%	+1.2%	+2.3%
World Total	+2.6%	+3.8%	+6.2%	+5.1%	1.1%	+3.2%	+3.2%	0.0%	+2.3%	+1.1%	1.2%

According to TED, global productivity will only rise 1.2% this year or half the level for the period 2011 – 2019. In mature economies that will be only 0.4%. In the USA home to deep learning, productivity will actually fall by 0.7% but this should be taken with a pinch of salt as official labour statistics are highly inflated by amongst other things the assumption of new business start ups which assumes the US economy is currently in an expansionary phase when the opposite is the case. TED assumes no improvement in 2024.

The immediate beneficiary of deep learning will be chip manufacturers such as Nvidia and the Utility companies because deep learning servers consume prodigious amounts of electricity. The outlook for Microsoft and Google is less clear cut despite what is occurring on Wall Street. Meta may have muddied the field for them with its open source generative model called *AITemplate.* Meta followed this up by releasing a large language model called *LLaMA*, and whereas Google and Microsoft increasingly hide the

mathematical values underpinning their system, Meta released them or as they are known, they released the "weights" governing their learning process.

At least Google is worried about the competitive threat posed by Meta's free platform as a New York Times article reported: "*Some within Google have also wondered if open-sourcing A.I. technology may pose a competitive threat. In a memo this month, which was leaked on the online publication Semianalysis.com, a Google engineer warned colleagues that the rise of open-source software like LLaMA could cause Google and OpenAI to lose their lead in A.I.*" In the end the big earners for the AI companies are the corporate users of these systems. Given commercial confidentiality and the ability to tailor Meta's system to their own needs, bespoke in-house systems could emerge which more than compensate for the loss of Microsoft and Google's much larger data bases. For example while chatGPT3 had 175 billion parameters, by the launch of chatGTP4, that had grown to 1 trillion.

AI is not omnipotent.

One of my favourite and consequential recollections from the war in Vietnam was the episode of the twigs. The Vietcong had detected what looked like twigs falling from the sky. They collected some and sent them off to the USSR for analysis. The answer soon came back. They were people sniffers designed to detect concentrations of troops then transmit this information back to the USAF so they could be bombed by the B52s. Asked how they could get rid of this menace; the suggestion came back to hang buckets of piss around the forest as this would be mistaken for concentrations of Vietcong. It is said the USAF bombed many a bucket until the Yankees twigged they were being played. The moral of the story is this: if you understand your enemy's technology you can use it against them thereby turning a disadvantage into an advantage.

The same applies to AI systems. No doubt the spooks have restructured their intelligence services by reducing their reliance on spies on the street and in your neighbourhood in favour of employing indirect spying via the internet and cameras both stationery and mobile as in drones. This dependency on electronics creates a vulnerability should their systems be hacked or got round.

And AI can be fooled. It is not bullet-proof. Nature the science magazine has an article showing how humans can fool deep learning programs by subtly varying the image which the system has been trained on. For more details view the actual experiment linked here. All it requires is inside information as to the structure of the algorithm. And in any revolutionary situation inside knowledge will be forthcoming because a revolutionary period is based on mass action, and mass activity always draws in a wide range of elements in society both within and without the working class. It tends to test loyalties to breaking point, which is why it is riddled with defections from the state apparatuses which the movement would encourage.

Conclusion.

Private property is a minefield, buffeted by conflicting interests and standing on unstable soil. It has to be redefined time and again in periods of transformative change, and no change is bigger than the one posed by 'AI'. This period of upheaval shines a light on the parties jockeying for position to maximise their return from this new technology, in particular Big Tech seeking to manipulate legislation for their own benefit.

Brian Green, 7th June 2023.