WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Decentralized Governance of AI Agent

Siddharth Kewat[1],DR. Krishan Kant Lavania[2], DR. VISHAL SHRIVASTAVA[3], DR. AKHIL PANDEY[4], DR. VIBHAKAR PHATHAK[5]

B.Tech Scholar, Professor, Assistant Professor
Computer Science & Engineering Arya College of Engineering & I.T India, Jaipur (302028) siddhu245016@gmail.com, krishankantlavania@aryacollege.in, vishalshrivastava.cs@aryacollege.in , akhil@aryacollege.in, vibhakar@aryacollege.in

ABSTRACT—

Artificial intelligence (AI) is to a great extent accountable for revolutionizing healthcare, finance, transportation, and governance in contemporary society. Ethical adherence, accountability, and justice are coming into question as AI systems become more autonomous and capable of making choices independently without the involvement of human beings. Companies, governments, and regulatory institutions are in command of most of the conventional AI governance frameworks. These frameworks often fail to offer transparency, discrimination, and representation despite offering governance. Centralized AI regulation can result in hardto-audit decisions, which is problematic from an ethical and legal standpoint. These models also introduce a point of failure, and therefore, AI systems are vulnerable to security threats, abuse, and possible manipulation by influential actors. It is increasingly clear that an AI governance system that is more participatory, accountable, and transparent is necessary. Rather than depending on one authority, decentralized governance is an attractive option because it decentralizes decision-making power. Blockchain, smart contracts, and Decentralized Autonomous Organizations (DAOs) are explored in this research to determine how they can be utilized to create a transparent, fair, and secure governance system for AI agents. An unalterable ledger developed by blockchain technology keeps a record of all AI activities, such as model changes, decision-making processes, and training data updates. network of validators, not a single central entity This makes it so that AI systems cannot be modified and can still be verified. AI agent decisions can be audited, verified, and reviewed by a distributed network of validators instead of a centralized one by combining AI with blockchain-based governance. This makes it impossible for AI systems to be modified and can still be verified. AI agent decisions can be audited, verified, and reviewed by a distributed network of validators instead of a centralized one by combining AI with blockchain-based governance.

1. Introduction

Artificial Intelligence or AI is fast emerging as a force for revolution, changing medicine, finance, transport, and government. From recruitment and lending to driverless cars and predictive policing, AI-driven decision-making is shaping choices more and more. But as the autonomy and sophistication of AI systems
increased, so did debate about how to regulate, ethically evaluate, and hold them accountable. Corporations, governments, or regulatory authorities hold power in traditional AI governance frameworks. Although these bodies offer control, centralized governance has a number of disadvantages, such as transparency deficiency, biased choice, security threats, and solitary points of failure. The existing governance frameworks tend to restrict
stakeholder engagement, making AI systems inaccessible and prone to malicious use, manipulation of facts, and ethical violations.

One of the central issues in centralized AI regulation is trust. AI systems tend to be trained on confidential datasets that have little outside inspection, and so it's hard to know if they're fair and reliable. AI systems often inherit biases from their training data in most situations, which then produce discriminatory results in employment screening, credit assessment, policing, and medical diagnosis. In addition, centralized control poses a threat of monopolization, as a few organizations or governments will have disproportionate control over AI deployment and policymaking. This can lead to threats of abuse, privacy infringement, and concentration of power in the hands of a selected few, contrary to democratic principles and equity. The transparency deficit of AI decision-making also makes it difficult for individuals to appeal or contest AI-made decisions, giving rise to the possibility of injustice and public mistrust of AI technologies.
Give your text character.

- **The Emergence of AI and What It Means for Us Regular Folks**

So, we're living in a time where AI is like the cool new kid on the block, right? This thing called artificial intelligence is popping up everywhere, from the robotic friends who help out in hospitals to the clever gadgets that think they know which movie we will like next. It's changing how we live, learn, play, and even handle our money. Basically, AI can make stuff work better, help us make smarter choices, and just generally beef up what we can do as humans.

But, just like with anything that's super-powerful, there's a bit of a worry club forming around it. Because these artificial intelligence (AI) devices' decision-making processes can be a little mysterious, some people wonder if they're being fair and safe. Plus, they're starting to have a say in how society ticks, and we want to make sure they're not throwing the whole system off balance. That's why we need some serious ground rules for AI, a kind of bouncer for the digital world to keep things in check. So, while AI is like the cool gizmo that can totally change the game, we've got to make sure it's playing nice and not pulling any funny business. That means setting up some guidelines so we can all get along and keep the future looking bright and shiny. Current Governance Hiccups with AI

So, we've got these AI governance situations that are usually like the boss of the whole show, with all the power in the hands of one big cheese, like a government, some company, or a bunch of rulemakers. But here's the thing: when everything depends on the one person in charge, problems are bound to arise. First of all, it's possible that things will become rather secretive, and not everyone will be privy to the truth, which is not cool. Then there's the possibility that the big cheese will start to believe they own the place, resulting in unjust decisions that could cause havoc for everyone else. And let's not forget that when you put all your eggs in one basket, it's super easy for someone to come along and knock that basket over, you know, like with hacking or playing dirty.

And here's the kicker: when it comes down to what the big guys want, the little guys and what they need are forgotten. Because it is not hearing from everyone, this indicates that the AI we are developing might not really get what we all want or need from it. Additionally, if it begins to make decisions that appear dubious or simply incorrect, we might all begin to question whether or not we can

truly entrust it with our lives. This way, everyone can control what our AI friends do and ensure that they behave nicely with everyone. It's all about keeping it real and making sure AI doesn't go all "Skynet" on us. We need it to be as fair and trustworthy as a golden retriever, not as sneaky as a raccoon in the kitchen at night.

2. Methodology

A number of steps are included in the research plan to determine how to manage AI agents without a boss: reading other people's writings, developing a strategy for the new system, experimenting with technology, evaluating its performance, and finally providing some guidance. To begin, we will examine a number of articles and other materials to gain an understanding of AI governance, blockchain, distributed ledgers, and smart contracts. This will help us spot any problems with how AI is currently run and give us a solid base to build a system where no one's the boss of the AI. Next, we're going to sketch out this new way of managing AI that's all about transparency and keeping everyone in the loop. We'll think about who's important, how to set up the rules, and how to use blockchain to make sure the AI plays fair. And we're not just going to stop at that, we're going to use cool tech like Ethereum to make it all happen automatically.

After that, we'll build this thing ourselves to make sure it works well. We will use a decentralized autonomous organization (DAO) to make sure that everyone can influence the AI's actions. After that, we are going to put it in real-world situations like smart doctors and self-driving cars to see how well it can make decisions without a human boss.

After we have finished that, we will carefully examine its transparency, fairness to all parties, userfriendliness, and security. This is like giving the system a check-up to make sure it can handle the real world.

In the end, we will record what we have learned and what we believe needs to be done next. This research will be like a road trip through various fair, open, and intelligent ways to manage AI, and at the end, we will have some cool ideas for other people to play with.

3. Defining the Ethos of AI Agents

The ethos of AI buddies, you know, the heart of what makes them tick in the right way, has gotta come from a set of core beliefs that guide how they behave and make choices. First and foremost, we must ensure that these smarty pants are made to always prioritize humans, safeguard us, and uphold our rights. They gotta be fair, not playing favorites, and definitely not being jerks to anyone, especially when dealing with all sorts of people. Secondly, they need to own up to their actions. In other words, if they make a mistake, we need to be able to understand why and hold them accountable. That means making it so that we can see what decisions they made and why. Similar to keeping a diary, but with algorithms. Third, these AI guys have got to be transparent, you know, like glass houses. And let's talk privacy, right? These AI systems need to treat our private information as if it were top-secret. They should totally be into saving the planet and helping everyone get along. And let's not forget that AI is there to help us, not to take over. They should be like Robin to our Batman, you know, helping us make better decisions without taking the wheel completely. And they should get us, respect our choices, and not be biased, because, let's face it, we're all different and unique snowflakes. Now, when it comes to being culturally sensitive, these AI systems got a get it. That way, when things get tricky, like in doctor offices or courtrooms, we're all on the same page. They ought to behave similarly to friends who constantly improve themselves and grow from their mistakes. Therefore, we need to keep an eye on them to ensure that they are always adapting to the times and are not causing any unintended drama. Last but not least, we want our AI to behave like chameleons and blend in wherever they go. They've got to be flexible and able to handle different situations without losing their moral compass. This way, no matter where they are used, they will continue to behave ethically and with respect for others. Therefore, AI that is not only extremely intelligent but also extremely cool results when we combine all of these factors. They are kind, considerate, and always looking out for the greater good, like the kind of friend that everyone wants. Isn't that what we want as well? AI that's not just good with tech stuff, but good for people and the planet too.

4. How the Ethos of AI Agents Informs risk-Based Regulation

The ethos of AI bots is super important in setting the vibe for how we regulate them, especially when we're talking about playing it safe and being fair. All that matters is establishing guidelines to ensure that these bots act ethically, such as not favouring anyone or acting as though they are superior to us.

Fairness is essential first and foremost. We need to make sure that these AI systems don't go around being biased or treating people differently based on stuff like race or if they're a boy or girl, or how much money they have. Regulations should totally keep an eye on this, so if an AI bot makes a dodgy decision, we know who's got to sort it out. In the AI ethos, accountability is another important component.

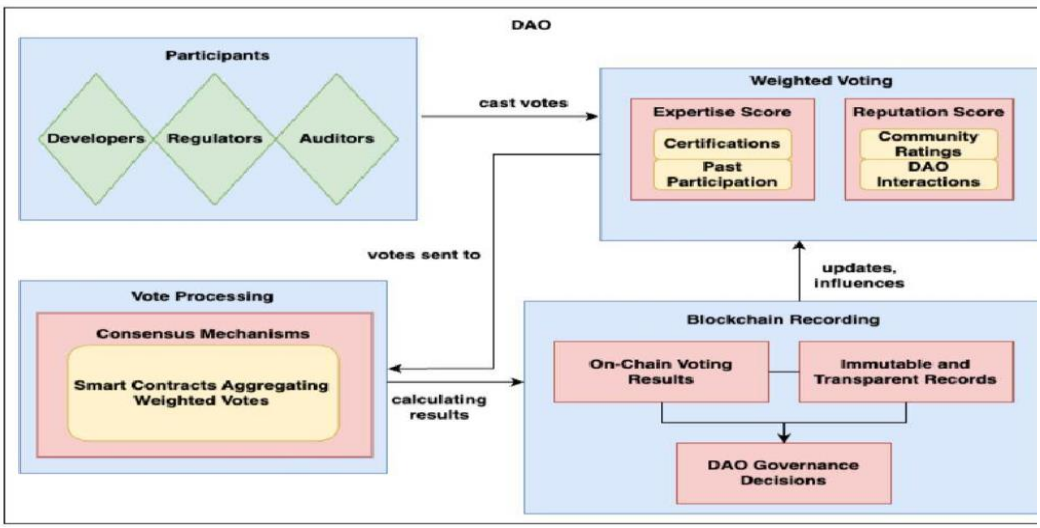
Transparency is comparable to the "glass house" rule in AI. So that we are aware of what's going on, we would like to be able to observe their decision-making process. Regulations should enable AI bots to clearly explain their choices to us. This helps us build trust and prevents us from becoming suspicious of their actions when it comes to crucial matters like healthcare or who is breaking the law. And let's not forget, AI should be all about helping us out, not taking over. That's what we mean by human-centric values. It's like having a really smart assistant that actually cares about what we think and feel. Regulations need to keep AI in check so they don't start making big decisions without asking us first, especially in places where our privacy is at stake.

Now, sustainability and thinking long-term is the cherry on top of the AI ethos sundae. Not only should these systems be made to be environmentally friendly in terms of how much power they use, but also in terms of how they help us deal with big problems like climate change. Regulations should keep an eye on the future to make sure that artificial intelligence isn't just a temporary solution that makes problems worse than it solves. Lastly, AI should be like a chameon , changing and learning as it goes. The ethos needs to encourage them to keep getting smarter while still playing by the rules. Regulations ought to be adaptable so that we can modify them in response to emerging issues as AI develops. That way, we can be sure AI keeps doing the right thing, even when it's dealing with stuff we haven't seen before.

So, the ethos of AI bots is like their moral compass. It helps us make sure they're cool to be around and that they play nice with us humans. If we get it right, we can have a world where AI does amazing things for everyone without stepping on any toes. We can all support a future in which regulations based on this philosophy ensure our safety and happiness.

The ETHOS Model

The ETHOS Model is like a buddy system for AI agents. It's got some cool rules that keep them friendly and helpful. First, compassion. AI should be like those friends who totally get you. They must be able to sense your vibe, comprehend what you stand for, and assist you when you need it. That way, you'll trust them more and actually enjoy hanging out with them. Then there is openness. Imagine if you were unable to understand why your friend did something. You'd be confused, right? AI needs to be like that friend who's always upfront about what's going on, so you know why they made a decision. As a result, everyone feels like they are in the loop and the trust is maintained. Human- Centric Design, you know, it's kind of like making sure everyone chilling at your buddy's place is comfy. AI should totally be designed to lend us a hand and make life a breeze, without, like, getting all up in our business or causing any awkward moments. Basically, it should be our helpful pal, not some clumsy oaf tripping over our stuff. When the stakes are high, they should emphasize our capabilities and always keep in mind that we are in charge. Open Accountability is like your buddy owning up when they mess up. If an AI agent makes a mistake, it should be easy to figure out what happened and who's responsible. If something goes wrong, everyone will know how to fix it, and no one will be held accountable. Sustainability is about making sure AI doesn't trash the planet like some kind of unruly party guest. They've gotta be designed to be ecofriendly and play nice with Mother Earth, while still helping us achieve our big goals, like making sure everyone's got a fair shake in life and tackling those global challenges together. Lastly, Security. Nobody wants to break into their friend's house. That's why AI systems have to be like a fortress, keeping all the bad guys out and protecting our info. If they're gonna be a part of our lives, they gotta be tough against cyber baddies and have a plan if anything goes haywire. So, the ETHOS Model is just like teaching AI to be good pals: understand us, be honest, think of the greater good, and keep us safe. If they follow these rules, we'll all get along swimmingly in our techy future.



5. Technological foundation

The technological components that make up the ETHOS model for AI agents are like a cool toolkit that enables these smarty-pants computers to behave more like humans, albeit with better manners and a sense of morality. It's all about mixing together some fancy tech stuff with a good dose of human values, so they play nice with us and do their job without causing a ruckus. Here's what's in that toolkit:

1.AI and Machine Learning Goodies: These are the brainy algorithms that let AI learn from stuff, kind of like how you'd learn from your experiences, except without the awkward phase. But these algorithms have to be really careful not to get biased, like that one friend who thinks their favourite team can do no wrong.

2. AI can communicate with us without sounding like a machine reading from a script by employing NLP chit-chat. It's all about understanding feelings and context, like when you know your friend is joking because of their tone. To help them become skilled at it, we provide fancy models like GPT and BERT. 3. Giving AI the ability to talk, or XAI, or "Explainable AI," increases our level of trust in them. XAI is similar to giving AI the ability to talk because it allows AI to make decisions that appear odd at times. It's similar to asking a student to demonstrate their work in math class.

4. Fairness and Bias Busters: AI can sometimes pick up on the wrong cues and end up being a bit biased, like thinking all dogs are friendly just because you've met a few nice ones. They are able to see the world more as we do, without prejudice, thanks to these methods.

5. Security and Privacy Gear: Think of this as AI's bodyguard, keeping all the sensitive stuff safe from digital peeping Toms and making sure they play by the rules, like not telling your secrets or letting anyone mess with their head.

6. Decision-Making Autopilots: These are AI systems that make choices without asking for a human's opinion. They have objectives they want to achieve, but they also have to be fair and keep everyone safe, just like a good game referee would.

7. Party Games Using Blockchain and Ledger: This technology is like having a diary that no one can change. It records everything an AI does so we can verify that they are not engaging in shady behaviour.

8. Technology for Human-AI Bonding: The goal here is to make sure AI talks to us like we're people, not robots. Think chatbots, voice assistants, and those cool AI buddies that totally get you. It's like chattin' with a pal who actually hears what you're sayin'.

9. we're talkin' about the AI Rulebook. This is like the set of rules you'd have for anyone hangin' around your place or workin' with you. It's super important so these clever little helpers don't go all wild, break the law, or accidentally make someone sad.

And number ten, Data Security, baby! You know how you don't want random folks snoopin' through your stuff? Well, AI's gotta be like the ultimate respectful guest and handle our personal info with kid gloves. It's all about keeping our digital lives as private as we keep our diaries. The AI Playbook: So, think of AI like having a bunch of rules at home or the office, you know? It's like these guidelines that tell the techy brainiacs what's cool and what's not. Super important to keep 'em from going bonkers, breaking laws, or making us feel bad, right? 10. Data Cleanup: Just like we don't want peeps rummaging through our personal stuff, AI's gotta be careful with our data too. It's all about respect and keeping things on the down-low. This is about keeping data clean, organized, and used only in ways that we're cool with. It's all part of being a responsible digital citizen.

So, that's the lowdown on the ETHOS model's techie side. It all boils down to making sure that AI is just as friendly, fair, and reliable as the average person, but with a lot more computational power.

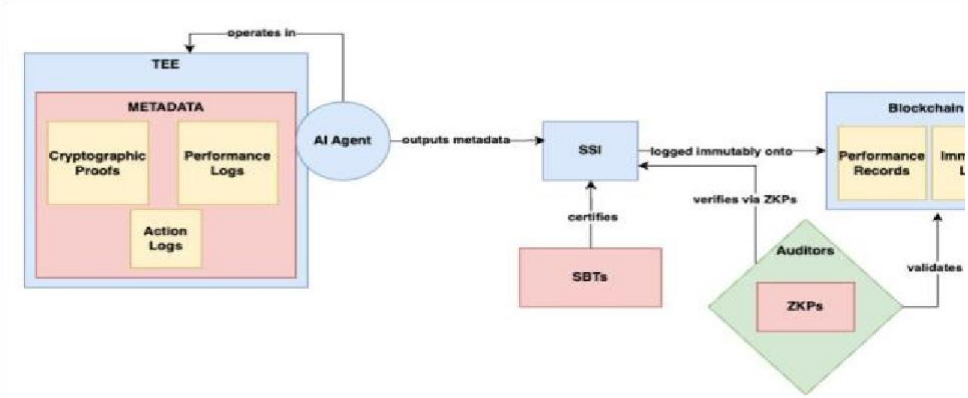
6. 5.3 Identity Management of AI Agents on ETHOS

In the world of decentralized ETHOS, identity management is of the utmost significance for AI. Giving AI friends their own IDs is similar, but there is no big boss in the middle. ETHOS gives each AI a unique, tamper-proof ID through the use of this cool blockchain technology. Self-sovereign identities, or SSIs, are the fancy names for these identification cards. What this means is that the AI can be in charge of their own info, like a digital boss. They get these credentials from trustworthy places to show they're cool to hang out with and they play by the rules. It's similar to having a VIP pass that lets people know you're real without showing off your private items. The best part, too? These IDs are like a universal language. AI can unwind in any platform without incident, adhering to the local regulations and disclosing their personal information. If an AI starts acting shady, the system has a way to put a stop to it without breaking the law. ETHOS keeps

tabs on everything with these nifty audit trails. It's on innovative mechanisms for resolving disputes like having a digital diary for each AI that nobody effectively and equitably. By harnessing the power can delete or change. This makes sure that everyone of blockchain technology and integrating innovative plays fair and square, and if an AI does something economic models, decentralized justice systems can they shouldn't, it's easy to catch 'em. So, in the end, adapt to the complexities of AI-driven societies**,**

ETHOS is like the cool school for AI where ensuring that governance frameworks remain everyone has their own ID, they can play nice with resilient, inclusive, and ethically sound. others, and nobody can pull a fast one without getting caught. It's all about trust, keeping it real, and

7. Decentralized Justice



are already in place. We want everyone to get a fair shake without throwing the baby out with the bathwater, you know? Finding that sweet spot where everyone wins is In a world where AI agents will have unprecedented everything.
influence over critical decisions, ensuring fairness, accountability, and transparency in their governance demands a justice system as dynamic as the technologies it seeks to regulate. Towards this end, the ETHOS framework offers a decentralized

8. Discussion

sticking to the rules. And it does this with some
fancy tech magic called blockchain, smart contracts, What's better? It's supposed to cut down on the whole and cryptographic handshakes. Basically, it's the corruption deal, keep costs low, and let everyone in on the future of AI, where they can do their own thing legal action because records can't be messed with. without stepping on toes. Additionally, it has an automated dispute process running, which sounds like a win-win situation. The kicker, however, is that there are still a few questions to be answered. For example, how can we persuade prominent legal figures to embrace this decentralized concept? And how can we ensure that these systems can efficiently handle a large number of cases? Oh, and there's the whole "fairness" deal with AI being in charge of decisions. We must prevent the robots from playing favourites, correct? Additionally, it is important to safeguard our legal identities. That's a biggie. So while this decentralized justice stuff could be the next big thing, we've got to tread carefully and make sure it plays nice with the laws that

9. Conclusion

approach to justice, designed to address the unique
challenges of AI governance while upholding ethical and legal principles in an increasingly Our main thing we're adding to the conversation here automated society. Decentralized justice is built is this cool concept called ETHOS, which stands for upon digital courts that leverage blockchain Ethical Technology and Holistic Oversight System. technology as the cornerstone for ensuring
transparency, accountability, and fairness in The ETHOS deal is all about using some seriously resolving disputes in the digital age. That is, through neat tech like blockchain, smart contracts, DAOs, game theory and mechanism design (Ast and SSI, and SBTs to keep track of AI guys worldwide. Deffains, 2021), settling disputes becomes a matter It's like a global Rolodex for robots, but way more of designing judicial architectures that leverage high-tech and secure. What's cool is that it doesn't economic incentives to crowdsource jurors, just slap a label on them and call it a day; it actually enabling peer-driven judicial decisions facilitated by understands that AI can get into all sorts of jams and smart contracts (Aouidef et al., 2021). These needs different kinds of supervision based on how mechanisms play a vital role in addressing disputes risky they are—from "Oh no, don't let it near the arising from the increasing autonomy and buttons!" to "It's all good, it's just playing with data." complexity of AI agents setting the stage for broader questions of legal liability. In the coming age of AI By mixing in some philosophy with the nitty-gritty agents, leveraging decentralized justice mechanisms of how these AI systems work, ETHOS tries to will be critical in its implementation relies heavily figure out how to balance giving them freedom to do
their thing with keeping them from going full Skynet on us. It's like a smart parent who lets their kid grow but keeps tabs without being too overbearing.

This system is supposed to be the be-all and end-all for making sure AI plays nice with everyone, and it does all the boring stuff like checking their homework, making sure they follow the rules, and keeping everyone in the loop—all while being super flexible and adaptable. But here's the catch: it's got to work in the real world, so we've got to make sure it gets used by everyone and gets tested like crazy before we can say it's ready to tackle the AI future head-on. So, it's a big deal if it actually takes off and everyone's happy with it.

10 . Acknowledgment

We would like to thank everyone who helped us with this research on decentralized justice. First and foremost, I would like to express my sincere gratitude to all of the academics, researchers, and technologists involved in legal matters, AI, and blockchain. You guys are the real MVPs here. Without your cool ideas and hard work, we wouldn't have had much to go on. You have really helped us comprehend the concept of decentralized governance and its potential to alter the justice system. We've got to give a shout-out to our mentors, professors, and the schools we're part of too. Additionally, the open-source and decentralized tech communities should not be overlooked. You all rock! Your teamwork and all the cool stuff you're building with blockchain, smart contracts, and those fancy dispute thingamajigs have totally blown our minds. We're here to support you as you open up a whole new world of legal technology. Last but not least, we've gotta thank our buddies, work pals, and family. We appreciate your constant support and encouragement. We have been driven to do our best by your support, which has meant everything. We hope this research we've done can actually make a difference and help create a legal system that's more see-through, fair, and easy for everyone to use. Cheers to all of you!

11. REFERENCES

- Satoshi's Bitcoin thing from 2008: <https://bitcoin.org/bitcoin.pdf>
- Vitalik's Ethereum brainchild from 2013: <https://ethereum.org/en/whitepaper>
- Lessig's deep dive into internet laws from way back in 1999: "Code and Other Laws of Cyberspace"
- Wright and De Filippi's look at blockchain and the law in 2015: "Decentralized Blockchain Technology and the Rise of Lex Cryptographia"
- The Kleros peeps' guide to decentralized justice in 2021: <https://kleros.io>
- The Tapscott siblings' take on blockchain changing everything in 2016: "Blockchain Revolution"
- Katsh and Rabinovich-Einy's digital justice book from 2017: "Digital Justice: Technology and the Internet of Disputes"
- And Szabo's 1997 paper on trust and networks: <https://firstmonday.org/ojs/index.php/fm/art icle/view/5472/4684>
- De Filippi and Hassan's 2018 article turning code into law: "Blockchain Technology as a Regulatory Technology: From Code is Law to Law is Code"
- And DAOstack's guide to making decisions in the Web3 world: <https://daostack.io>