"Artificial Intelligence and the Political Philosophy of the Future – Smarter Planet or Wiser Earth?'
Text to read – Gray Cox, Draft 8/24/22

"AI and the Political Philosophy of the Future: Smarter Planet or Wiser Earth?"

Keynote Presentation for the 12th NIC.br Annual Workshop on Survey Methodology
São Paulo, Brazil, August 29th, 2022

Gray Cox

Professor at College of the Atlantic
Clerk of the Quaker Institute for the Future
gray@coa.edu, #207-460-1163

S 1

I am very grateful for the invitation to share this research on Artificial Intelligence and the political philosophy of the future.

As a teenager, learning to do construction work, I got some advice from old timer watching me run around, gruntin, scurrying, lugging cinder blocks, shovelling sand. "Work smarter, not harder." It was good advice.

Nowadays I see many of us scurrying around in other ways, losing track of the larger frames of our lives – suffering burnout, addiction, divorce, health collapse as individuals and, other analogous crises as communities and countries.

What kind of thinking might be most helpful as our world becomes increasingly dominated by "smart" cars, farms, factories, schools, cities – smart battlefields? In this new context, we may need to follow another kind of advice: "Live wiser, not just smarter."

The overall aim of this talk is to explore what this might mean.

S2

In this talk I will try to:

1. Frame key problems for AI and our future

2. Distinguish Two kinds of reasoning -- monological inference vs. dialogical negotiation – to help us address those problems
3. Describe Two models of AI – the Turing Machine vs. the Turing Child
4. Explain 7 Strategies for developing Turing Child systems in programming and collaboratively solving our problems

S3

Sometimes slowing down and taking a deep breathhelps the work go better. I hope you won't mind if I calm a bit by sharing a song with you to help set the mood:

I'm gonna slow right down,
so I can get there sooner.
I'm gonna slow right down,
so I can get there today.
I'm gonna slow right down,
maybe even come to a full stop.
Maybe if I come to a full stop
I'm gonna get there right away.

S4

Part I, framing the problems . . .

S5

Gandhi once commented:
"Civilization is not an incurable disease. But we should always remember that the English people are currently afflicted by it."

What might this mean?

S6

Our global civilization is structured by ways of reasoning in economics, governance, technology and morality that threaten our species with:

1. ecological collapse,
2. pervasive injustice & the threat of mutually assured destruction,
3. domination by super-human machine intelligence and/or foolishness
4. moral relativism and the annihilation of meaning for human life

Imagine an alien anthropologist from Alpha Centauri arriving on Earth and observing all this. Her first note home to her advisor?

"A species which imposes such radical existential threats upon itself --- what are they thinking????!!"

The alien graduate advisor's likely reply might be:

"Clearly their dominant reasoning strategies are, in a profound sense, irrational. The central research question is: HOW are they thinking?"

S7

I want to suggest our dominant idea about how to think best is to try to think ever "smarter". This is part of a project since the 18th Century Enlightenment and the Industrial revolution to develop

what IBM has called, a “Smarter Planet”. A planet where everything and everyone is instrumented, interconnected and intelligent.

S8 S9 S10
Theorists like Ray Kurzweil and Max Tegmark have laid out visions of how all this may culminate in some form of Artificial Super Intelligence.

S11

But in considering this dream of Artificial Superintelligence, perhaps we need to reflect more. What exactly is it? Intelligence?

S12

Let me offer a proposal for our purposes today:

Intelligence is the ability to sustain and/or enhance one or more values in various contexts over time.

S13
Note some key features:

1. Intelligence, in this sense, is **guided by values**. We can only distinguish more vs less intelligent behaviors if we care about consequences – In a world without values, there are no wrong answers and no smarter methods.
2. Intelligence **reshapes or adapts the self and/or the world** to reflect those values.
3. It can take **many forms** -- calculating a solution, negotiating an agreement, writing a melody, constructing a piece of furniture, sharing an intimate feeling, cooking a new dish, keeping warm, nurturing an offspring . . .
4. In this sense, organisms and biological communities may exhibit intelligence and so may machines and other systems – **“intelligence” in this sense does not require consciousness**.
5. **“Intelligence”** may be partial and limited, falling short of a **wisdom** that responds appropriately to the full range of values we should hold in our lived context.

S14

Contrast Intelligence with ***Wisdom***
which we might tentatively define as:

“***systematic intelligence that responds appropriately to the full range of values we should hold in the context in which we live***.”
In that sense wisdom is human ecological.

Unfortunately many who aim at high levels of intelligence often focus on only one or a few relevant values – un-wisely.

S15

Artificial Intelligence is:

- Created by **"artifice"–** a design process at least initially, in part, guided by explicit intentions
- Typically **silicon** based but **need not be**
- Traditionally programmed by a person or team but **can be designed to use evolutionary processes,** for example**, to program itself.**

Today I will use this intentionally very broad definition of AI that includes everything from the minimal intelligence of heating thermostat controlled to the most advanced forms of machine learning systems like GPT-3 and Wu Dao 2.0

S16

It is useful to distinguish varieties of Artificial Intelligence (AI) as:
Narrow AI
Artificial General Intelligence (AGI)
Artificial Super Intelligence (ASI)

S17

Our civilization is pursuing two projects, the ever smarter planet, and the Artificial Super Intelligence to run it.

A key danger of the Smarter Planet project is that it can turn our life systems over to managers that are "smarter" without being wiser. They may maximize one or a few values very efficiently, but leave out others of vital importance. Think of the many functions farmlands serve in ecosystems. If a "smart system" manages them solely to maximize production of tons of soy beans per hectare, think of how it will foolishly ignore the vital ecological services of hydrology, wildlife, carbon sinks, and conservation of land race seeds, – and vital social functions like providing jobs, homes, green spaces and community identity.

A key danger in pursuing Artificial Superintelligence is sometimes referred to as "The Friendly AI" problem, often assumed to be simply getting AI's values to align with ours. But there are two distinct problems here:

First, we don't want an ASI to simply align its values with its creator or owner – bad people could use it for very bad purposes. So there is the challenge is to create AI that promotes what is good and right.

But then, second, how well do our own values align with what is good and right? If they don't, then the ASI working for the good will be critical of us. How will we merit its friendship?

S 18

To address these challenges, I want to distinguish two kinds of reasoning: monological inference vs. dialogical negotiation

S19

Part 2.) A. Reasoning as Monological Inference employing algorithms

The first systematic articulation of inference in a Formal Logic was provided by Aristotle's theory of categorical syllogism with its
set of algorithmic rules used to generate categorical conclusions from categorical premises. For instance, the rule that:

If "All A are B" ***and*** "C is A", then "C is B".

To apply it in a rational argument you simply need to input a set of premises like, for example:

All men (A) are mortal (B).
and Socrates (C) is a man (A).

and then *run the algorithm and get the conclusion that:*

Socrates (C) is mortal (B).

S20

There are many ways algorithmic thinking has been embedded in our civilization.
With Euclid's Elements of Geometry and then Newton's Physics, very sophisticated and powerful theories of space and of the laws of motion were developed and were so impressive that, for many people, they were taken to be the paradigms not only of scientific achievement but of rationality itself.
This inferential model of reason is monological in the sense that it starts from one point of view, one set of premises, and draws conclusions. And it can be all carried out by a single individual – or machine. Given the right data and axioms, a Newtonian can work wonders predicting, for instance, how to shoot a rocket to the Moon.

S21

In the 18th Century, Jeremy Benthan and Immanuel Kant adopted this same core conception of rationality as their model for thinking about ethics. They sought one or a few principles to enable them, like Newton, to provide "laws of moral motion" to create an ethical rocket science they could use for a set of premises about the data of the world and the axioms of ultimate value to then, following logical algorithms, infer conclusions about how to act.

S22

For Utilitarians – like Bentham, the fundamental axiom of ultimate value was the Greatest Happiness Principle:

Always choose that action that will yield the greatest net happiness to all concerned!

S23

For the Duty based ethics of Kant's Categorial Imperative, the fundamental axiom could be expressed as:

Act only according to that policy that you can, rationally, at the same time, will as a universal law!

or, alternatively:

Second: Never act according to a policy in which you treat other rational creatures as means only (mere things) but rather, always with **respect** as **ends in themselves** (persons)!

In Mainstream Western Moral theory, at least as practiced in English speaking countries where much of current work on ethics in AI is done, the dominant focus is on alternative variations of the Utilitarian and Kantian approaches.

Sometimes the two yield essentially the same practical conclusions. But in many important cases they don't. Researchers and students are faced with dilemmas –

S24

Which approach to choose?

In the US, a dominant pedagogical approach is to focus, in university classes, on ethical situations like the Trolley Car Dilemma -- for two purposes: 1. To force students to clarify their own intuitions or prejudices about which ethical principle they believe is more fundamental and 2. To give students practice in the kind of moral reasoning both Bentham and Kant assumed was appropriate in ethics, namely, monological processes of inferences using algorithms to go from premises to conclusions.

There is a fascinating documentation of this pedagogy in a youtube video of an exemplary teacher, Michael Sandel of Harvard University.

S25

He has students imagine a Trolley car rolling unstoppably down a track towards five unaware people who it will run over and kill -- unless . . . it's diverted to an alternate track. You can see the car and the people AND you have a switch that will shift the track and make the Trolley go left and avoid those 4 people. BUT there is an innocent person standing on that other track will be killed as a result. The dilemma? Should you push the switch and kill the one to save the many?

He has students offer their views and draws out of them arguments that illustrate contrasting Utilitarian and Kantian analyses.

S26

For instance,
A Utilitarian might argue to sacrifice the one for the many and pull the switch.

A Kantian might hesitate, or even refuse to pull the switch because this might treat the one who is sacrificed as a mere thing, a means to save the others – and because she could not will it from the point of view of the person to be sacrificed.

Students struggle with this dilemma but most gravitate towards the utilitarian choice.

S27
Then Sandel presents a variation on the dilemma:
Imagine you are a doctor with five patients in your clinic, each needs a different organ for a lifesaving transplant.
And you have a healthy patient asleep
in the waiting room . . . Is this not the same dilemma? You could anesthetize the healthy patient, harvest his organs and save the other four.
What should you do
As a rational Utilitarian?
As a Kantian?
Sandel asks for a show of hands as to how many would sacrifice the one for the many.

The students audibly groan, horrified that as rational utilitiarians they seem obliged to sacrifice the innocent sleeper. But then one student up in the balcony raises his hand. Sandel is excited. A diehard utilitarian? But the student offers a very interesting comment: "I wouldn't sacrifice the healthy person. I would get sick person dying anyway anyway and to sacrifice themself for organs to save the other four."

When they hear this suggestion, the students burst into applause. This is great! A way out of this horrible dilemma!!!!
With practiced calm, Sandel lets the excitement die down and then turns to the class and says: "Well, that's a good idea. In fact, that's a great idea . . . except for the fact that it completely ignores the whole point of the philosophical example." Then he abruptly drops the discussion and turns to another case. In doing this as a university professor of philosophy he follows common practice – reinforcing the idea that ethics is about learning to make hard choices in the face of dilemmas -- and the idea that the rational way to make such choices is to pick your moral principles and stick to them, using monological processes of inference.

S28
But notice:
In real life, the student with the third option is just the kind of innovative thinker we would want on our team. We want folks like him in the dialogue brainstorming other creative options – like polling the terminally ill to see who might have a motive for making such a sacrifice. The search

for new ways of framing options available – and people's underlying interests -- can often provide "win/win" outcomes by "increasing the size of the pie" or even provide outcomes that stop framing the situation as a conflict with winners and see it instead as a shared problem participants are collaborating on.

Approaching ethics this way, we would want to make use of an alternative model of rationality.

S29
Part 2.) B – Reasoning as Dialogical Negotiation following guiding strategies

S30
Instead of algorithmic rules to follow, Dialogical Reasoning is structured by **strategies that guide**. They invite and suggest methods of observation, discernment, search and creative invention. Instead of an inference from premises to conclusion, as in formal logic, the process of reasoning would be of this basic form:
Step A. Encountering a difference with Other(s) →
Step B. pursue strategies of negotiation/problem solving in dialogue →
Step C. . . . till reaching genuine, voluntary agreement.

The Harvard Negotiation Project's *Getting to Yes* proposes guiding strategies like:

1. Multiply Options! (That is, if you face a dilemma, look for alternatives!)
2. Focus on interests behind positions! (Learn more about what the other person really wants so you can explore alternative ways to satisfy their concerns.)
3. Separate the people issues from the engineering problems! (Sometimes a relationship cannot be healed with any amount of money or material goods – it needs an apology or a public acknowledgement.)
4. Look for objective criteria! Seek out procedures and standards that are independent of your individual wills which you might agree are reasonable bases for arriving at sustainable, just, satisfying agreements based on emergent truths.

S31
In the last fifty years, research on negotiation and conflict transformation has yielded detailed accounts of these strategies and a host of others that help parties "get to Yes", engage in group problem solving, community based collaboration, mediation, dispute resolution, conflict transformation and peacemaking.

And studies of dialogical reasoning have
spread to a wide variety of other fields . . .

S32
Getting to Yes is a classic manual on negotiation strategies used in North American and international settings.
Lederach's Preparing for Peace draws on traditions of ethnography, appropriate technology, as well as Paulo Freire's community based approaches to collaborative learning to demonstrate methods for studying and improving practices in different cultures for negotiation, conflict

resolution and other forms of dialogical reasoning. Ramsbotham, Woodhouse and Miall provide an excellent survey of a host of these traditions in their Contemporary Conflict Resolution.

S33
In different ways, these research traditions aim to shift our civilization – following Gandhi.

We live in a culture in which peace is obscured, defined in terms of what it is not, and as a state rather than an activity like its "opposite", war.

This is because our culture is dominated by practices that assume conflict is essential to life.

S34
The result is a Culture of conflict, in which the
core metaphor for life is:
Two Islanders and one coconut . . . and they both want it. On this view everything else -- economics, politics, religion, . . . – just adds more islanders and more kinds of coconuts to a situation that ultimately reduces to conflict.

S35
Many of the studies of dialogical reasoning start with that conflict centered view and simply look for ways to reduce the violence used and move away from vengeance cycles of 'lose/lose" struggles to transform the goal from beating others into achieving "win/win" solutions.
Other practices of dialogical reasoning enrich the range of metaphors with life as shared problem solving or a dance.

S36
Or they may use the process of birth as a metaphor for life – birth is seen as a struggle in which lives may be at stake but there is no conflict. . . . but the pregnant woman and fetus are not trying to "win" against the other. Instead, both are sharing a process in which they are transformed as individuals into mother and child and enter a new set of community relationships.

S37
What are some distinctive features of these traditions of dialogical reasoning?
First, they understand the reasoning process as involving two or more real people with substantively different language, beliefs, and norms for starting points.

The challenge for these parties is to negotiate those differences and develop new language, practices and plans of action on which they can agree.

S38
Second, they commit to seeking genuine agreement through nonviolent practices of investigation and persuasion, without threats or coercion.

S39
Third, they use variations of the four basic strategies of: Multiply Options! Focus on interests behind positions! Separate the people issues from the engineering problems! Look for objective criteria!

Note that these take the form of open-ended imperatives that guide. **They are not algorithms.**

A fourth common feature is a shift away from the *Golden Rule which says*: Do unto others as you would have them do unto you!

As typically interpreted, that rule can provide an ethnocentric approach inviting colonialism and imposing our own preferences and values on others.

Instead, we are asked to start by inquiring into others interests and values and shift to the ***Rainbow Rule:*** Do unto others as **they** would have you do unto **them**!

S40

A fifth feature of these exemplars is that they understand the elements and aspects of the reasoning process in "emergentist" rather than "static" or "reductionist" ways. For them, the meaning and truth of sentences, the identities of the selves and communities stating them, and the social realities involved all emerge and grow or otherwise develop during the dynamic course of negotiation.

Many of the distinctive practices of rationality in these traditions focus, precisely, on methods for getting shared meanings to emerge in forms that express increasingly truer views of our options and become more agreeable for all.

In his "Experiments with Truth", Gandhi developed a kind of experimental method for discerning, demonstrating and defending emergent objective moral truths through practices of what he called non-violent "clinging to truth" or *"satyagraha"*.

S41

If there is time later we might explore how such methods can enable people to discern, demonstrate, and defend moral truths.

S42

Along with the systematic statistical studies of progress in their rates of success.

S43

But, to sum up, so far, I suggest we need to move

from a primary reliance on the 18th Century model of rationality as monological inference that makes us "smarter"

to a more inclusive 21st Century model that draws on monological reasoning to express individual voices but then seeks to resolve their conflicts through **more inclusive** forms of **dialogical rationality** that make us **wiser**

and help us deal not just with "complicated" problems like landing on the moon but also with **"complex" or "wicked" problems** like ending poverty – problems that involve multiple, divergent and incongruous perspectives and frames of meaning for understanding values, elements and dynamics.

S44
If we have more time at some point, I would be very interested in talking about how these two paradigms of reasoning are transforming the ways in which people understand economics, politics and morality. However, at this point, I would like to focus in on their relevance to Artificial Intelligence and the ways its technologies are transforming our world.

S45
In particular, I want to talk about Two models of AI – the Turing Machine vs. the Turing Child

S46
The work of Alan Turing (1912-1954) played a key role in the development of computer science in general and AI in particular. I want to focus on his classic paper,
"Computing Machinery and Intelligence" where he introduced the familiar Imitation Game as a way of operationally defining computer intelligence.

S47
The paper also developed a clear, explicit account of what is often referred to as the "Turing Machine"
-- the defining "Standard model" of the modern programmed, inferential, algorithmic computer

What is generally overlooked, however is that Turing also, in a final section of the paper, introduced a different, second basic conception of AI, the conception of what we could call the "Turing Child" – which he offered as a vision of a machine that learns through dialogue and socialization.

S48
In introducing his conception of what became the "standard model" of the "Turing Machine, Turing says:
"The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer. The **human computer** is supposed to be **following fixed rules**; he has **no authority to deviate from them in any detail**. We may suppose that these rules are supplied in a book, which is altered whenever he is put on to a new job. He has also an unlimited supply of paper on which he does his calculations."

S49
Some key elements and functions of the Machine involve: Input, Storage, Algorithms and Output

Turing was really an extraordinary thinker. After developing this Machine model at some length and describing how it might make progress in the future, in the closing section of the paper he introduces a really revolutionary, second model, the Computer as Child rather than machine. He notes:

S50

“In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components:
The initial state of the mind, say at birth,
The education to which it has been subjected,
Other experience, not to be described as education, to which it has been subjected.
Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child’s? If this were then subjected to an appropriate course of education one would obtain the adult brain.”

S51
It is important to stress that this second model is not a tool that is programmed by a user; it is a child that is educated in a community. Turing goes on to note that:
“It will not be possible to apply exactly the same teaching process to the machine as to a normal child. It will not, for instance, be provided with legs, so that it could not be asked to go out and fill the coal scuttle. Possibly it might not have eyes. But however well these deficiencies might be overcome by clever engineering, one could not send the creature to school without the other children making excessive fun of it. It must be given some tuition. We need not be too concerned about the legs, eyes, etc. The example of Miss Helen Keller shows that education can be take place provided that communication in both directions between teacher and pupil can take place by some means or others.”

S52
Note then some key features of a “Turing Child”:

The child machine will need to have a **body** and **engage in dialogical reasoning and interaction.**
The “programming” structuring such behavior will require kinds of interaction that are not monological reasoning or algorithmic calculations taking place in a formal “object” language. They will have to involve dialogue in which the **teacher and child machine** repeatedly **renegotiate the meanings** of terms and sentences.
They will also have to be able to move back and forth between the object language and the **meta-language** standpoints.

S53
In that regard, with a reference to Bertrand Russell’s introduction theory of types to avoid paradoxes of self-reference and infinite regress, Turing makes the following very revealing comment:

“The processes of inference used by the machine need not be such as would satisfy the most exacting logicians. There might, for instance, be no hierarchy of types. But this need not mean that type fallacies will occur, any more than we are bound to fall over unfenced cliffs. Suitable imperatives (expressed within the systems, not forming part of the rules of the system) such as ‘Do not use a class unless it is a subclass of one which has been mentioned by teacher’ can have a similar effect to ‘Do not go too near the edge.’”

S54

I would like to suggest that

We are reaching a critical moment in which obstacles to creating ”Turing Child” machines may be receding. It is the stage Max Tegmark describes as “Life 3.0” – with entities that can intentionally redesign their hardware and software.

We are, further, at a stage in which cutting edge textbooks in AI are reframing their core goals. For instance, Stuart Russell and Peter Norvig note, in the newest edition of their classic intro book:

S55

”Previously we defined the goal of AI as creating systems that try to **maximize expected utility**, where the specific utility information – the objective – is supplied by the human designers of the system. **Now we no longer assume that the objective is fixed and known** by the AI system; instead, the system may be uncertain about the true objectives of the humans on whose behalf it operates. It must learn what to maximize and must function appropriately even while uncertain about the objective.”

-- Stuart Russell and Peter Norvig***, Artificial Intelligence: A Modern Approach***, fourth edition, p. vii

S56

I want to suggest that passages like this are evidence of a growing awareness of the need to introduce dialogical approaches to reasoning in work on AI – not just in the ways the computers are structured but, perhaps even more, in the ways we understand the entire systems of society, technology and nature of which they are a part.

S57

This chart sums up key contrasts between the two models of AI and since the Turing Machine Model should be quite familiar let me just repeat the key features of the Turing Child model:

Two Approaches to AI	Turing Children

Style of reasoning	Multiperpectival, collaborative, dialogical
Process of reasoning as inference. Vs. negotiation	Uses guiding strategies to arrive at shared solutions or genuine, voluntary agreements
Starting point	Different points of view with different meanings ascribed to terms and different beliefs and rules

Process of reasoning	problem solving and conflict resolution in which any meaning, belief or value can be renegotiated
Goal of reasoning	Reach genuine voluntary agreements
Conception of truth	Cultivation of shared understanding of emergent objective reality

Status of reasoner	Not substrate independent, must be an embodied agent engaged in dialogue in open-ended contexts that include life worlds
Method of enhancing reasoning	Socializing the agent in lived contexts through parenting, teaching, play . . .

S58

NOTE: the difference between a Turing Machine and a Turing Child is not a matter of consciousness or using some kind of breakthrough technology like quantum computing.

It draws on a familiar basic process we all are born capable of and can learn and teach. We can improve our skills at it. We can incorporate them into our practices as individuals, communities – and programmers. And we can get computer systems to nudge us to use them more, use them better, and incorporate them in the reasoning processes of the machines themselves.

S59
Part 4.) Strategies for developing Turing Child approaches in programming and collaborative problem solving

S60
A key part of the shift involves thinking, explicitly, about how groups of people can incorporate dialogical methods in the decision processes that groups they go through when using computers.

This can start in extremely simple ways. For instance, I have written a program in SCRATCH which is designed to help children learn a bit about block coding and ethics by playing with a character called Ethel the Ethical Consultant Robot who helps them decide what to do in a Trolley Car kind of dilemma where they have to rescue one of two people from getting eaten by a Bear. The choice is presented as a dilemma using classic monological kinds of inference.

S61
BUT, near the end of the first round of the game, the kids playing are asked to list some ways in which they might be uncomfortable with their choice and the way the dilemma is presented. Ethel then invites them to list ways in which they might multiply their options, focus on underlying interests of people involved, and perhaps even change the kinds of criteria and calculations that are allowed in the game. – **and then** Ethel invites them to go into her programming and rewrite it to include those options – and change the game. In effect, she gives them a prompt to help her reason in new ways in the future.

S62
Variations on this same strategy can be done in working with adults in any context in which AI is being deployed in a community. For example, an organization developing software for reporting sexual assault can directly involve survivors in the ongoing redesign of the system using dialogical methods and incorporating elements of them into the group processes and structures as well as the program.

S63
Likewise, an organization like Consul can develop software for municipal governments using open source sites like GitHub that allow for version control. And it can adapt them to create ease of entry and interaction for community members who are not programmers.
They can be invited into dialogues to critique and revise the AI – and in doing so, they can provide exemplars of dialogical reasoning from which the AI itself may be able to learn.

S64
I would propose Seven Key Principles of Dialogical Approaches to AI/Human/Nature Systems:

1. First, in concepts, diagrams and practice, the projects should always be **framed as an AI/Human/Nature systems.** Intelligence is always an activity guided by values and concerns whose meanings are grounded in a holistic context. *Machines can make bits. Only a community can make a meaning.*

S65
So, in developing software for a farm or trucking company or restaurant or other part of a food system, it is essential to conceptualize the work not just as something an individual does as a programmer of a computer but more broadly as something an community does as a shared process of dialogue that includes people and other natural organisms as well as machines.

S66
Second:

The overall **goal** is to arrive at **genuine, voluntary agreements** – not to simply generate output -- genuine, voluntary mutual agreements between the AI, people and other natural organisms and ecological systems involved in the community engaged in the values and concerns at stake.
To be genuine, these agreements must be
understandable
based on consent
in a non-coercive context
and guided by emergent objective values

S67
3. The AI procedures need to **flag for review** the cases in which their data, algorithms, framing assumptions and/or outputs are especially questionable and need review by a human or by a representative group of humans and natural organisms from the larger community.

There should be ways for the machines to flag
e. g. sensors' margins of error, limiting features of training data for facial recognition, possible dangers in change of context for application, and high risk factors
machines should be able thus to initiate negotiations
and also advocate appropriately for values and concerns that call for systematic consideration

S68
4. The algorithms of the programs can be modified to conform to the agreements arrived at through **easily engaged meta-operations** that can be relatively easily accessed by other participants in the dialogical process.

through direct intervention by people
through reform of the machine's own programming
and do so in combination with principle #3 – e. g. flagging with phrases like "What should I be looking at here?" and "Do you have any idea why I seem to keep getting these two things confused?" or "I don't get it. Why isn't this one an X?"

S69
5. The dialogical interactions the AI engages in should be framed and guided by **principles of conflict resolution** as illustrated by Roger Fisher et. al.'s *Getting To Yes*, John Paul Lederach's *Preparing for Peace*, and other studies of negotiation and conflict transformation practices from around the world.

For example: focus on underlying interests, multiply options, "separate the people from the problem", Look for independent, objective criteria
flag problem points
generate a library of specific proposals as well as strategies

S70

Sixth: The AI should have a structure and **committed embodiment** that commits it to interests in the well-being of the community in which it is operating.

interdependent with the people and natural systems it engages with
tied irrevocably to physical machinery and power inputs that depend on the community for their maintenance -- should not exist merely as a cloud entity that is substrate independent
It can **"emigrate" or become exported AI capital only through genuine, voluntary agreements with the community** that created and maintained it up to that point

S71
Seventh, we should work to strengthen processes through which the AI/human/nature system can **discern tacit patterns** in the meanings that provide the context of its thought and action and make them explicit in spirit-led dialogue.

tacit patterns of **value as well as fact**
include both **physical** or material patterns but **also emergent formal and meta-structural patterns**
humans involved use reflection, meditation, "meeting for worship for discernment" & other methods to practice spirit-led communal discernment
also experiment drawing on the distinctive forms of intelligence offered by machines and by natural systems
"holding in the Light" not only natural systems like watersheds or forests but also the machines and artificial intelligence systems

S72
I look forward to conversation about how these 7 ideas for "AI as Collaborative Wisdom (CW)" might be applied in practice in your own work and now
turn the mike over to Joao for comments.
S73 [acknowledgements]

In summary: **Human Ecological Principles of a Collaborative Wisdom Approach to Dialogical Programming:**

1. framed as an AI/Human/Nature systems
2. goal is to arrive at genuine, voluntary agreements
3. AI procedures need to flag for review
4. easily engaged meta-operations
5. principles of conflict resolution
6. committed embodiment
7. discern tacit patterns in spirit-led dialogue

How might these 7 ideas for "AI as Collaborative Wisdom (CW)" be applied in practice in your own work?

Acknowledgements:
In working on these ideas I am very grateful for help from colleagues:
at the College of the Atlantic including, especially, Dave Feldman, Kyle Shank, Chris Petersen, Ken Cline, Davis Taylor, Suzanne Morse and my student Phileas Dazeley-Gaist,
at the Quaker Institute for the Future including, especially, Judy Lumb, Larry Jordan, Geoff Garver, Keith Helmuth, Charles Blanchard, and Sara Wolcott
at the Friends Association for Higher Education including, especially, Laura Rediehs and Don Smith
and at large including, especially, Ram Subramanian, Rich Hilliard, Brian Jennings, and Katie Wasserman
as well as support from the US National Science Foundation for research on differing cultural norms and their effects on the adaptation of technologies and the adoption of them.