Augmented Human “Intelligence”(?) in the new Hybrid and Mixed World: Autonomy and Roles of Intelligent Social Agents

Cristiano Castelfranchi

Institute for Cognitive Sciences and Technologies – Roma

WIC’16 Lipsia
We/You: creators of “autonomous” “intelligent” “social” Agents and hybrid societies and intelligences ....
We are not just building a **new technology** but a **new Socio-Cognitive-Technical System**, a new form of society, an anthropological revolution.

**We are social engineers.** Are we aware of?

I will focus on some problems and dangers of the Digital and WEB Revolution and of the **“mixed”** (virtual and physical) reality and **“hybrid”** society (natural and artificial intelligences) we will live in.
WE – AI & MAS community - are responsible for the introduction of “Agents” as “autonomous” (proactive, with initiative, with their own learning, reasoning, evolution, .. ) and “social”, cooperating with human by following true “norms” (but also – in case – violating them), and critically adopting our goals (not just “executing”), with over-help, critical-help, ....

And this was a correct and unavoidable solution, for a real “Intelligence” interacting with us and usable from humans.

And also for “science”: for modeling and understanding human intelligence and sociality.
“for POITING to some UNDERLYING THEORETICAL NOTION and ISSUE"
DEPENDENCE

AUTONOMY:

Pro-active: the Ag can anticipate us, takes the initiative; has its own information, and understanding, and learning, and abilities, ..

BUT: FOR WHAT & FROM WHICH POWER over him/it

Goal-Adoption is not “execution”, “obedience”

Goal-DELEGATION; RELIANCE: levels of delegation, adjustability, etc.

TRUST: components, degrees, foundation, ...
Goal-Adoption:

- **Spontaneous** (not requested!)
- **Over-Help** (beyond request)
- **Critical-Help** (even violating request)

but “for our good”

“tutelary” role
I’m not repented AT ALL, of modeling **ARTIFICIAL** intelligence and sociality on the contrary.. however

This obliges you **to become aware** of possible appropriation of your creations, of possible unacceptable uses of these instruments. There are **dangers in ICT and AI** and in particular in Autonomous Agent and Augmented/Hybrid Intelligence.

**Are we missing the control?**

**Not** of our Autonomous Agents, Robots, etc. but of **their** possible uses?

Are we ready for the **ANTHROPOLOGICAL REVOLUTION** grounded on Intell Technologies and artificial mixed society? Which also is **an economic, social, and political revolution**.
For the mass media, the main **PROBLEMS** are:

- Privacy
- **Security** (on WEB, ... on access ..)
- Hackers’ attacks
- Anthropomorphism
- **War and Artificial soldiers/arms**
- Ethics **inside** Artificial creatures and algorithms
For the mass media, the main **PROBLEMS** are:

- **War and Artificial soldiers/arms**

  Subra Suresh, Carnegie Mellon’s president, said *injecting ethical discussions into A.I. was necessary* as the technology advanced. While the idea of “Terminator” robots still seems far-fetched, the United States military is studying autonomous weapons that *could make killing decisions on their own* — a development that war planners think would be unwise.

Finally solved the problem of the poor general:
“General, your tank is a powerful vehicle
It smashes down forests and crushes a hundred men.
But it has one defect:
It needs a driver.

General, your bomber is powerful.
It flies faster than a storm and carries more than an elephant.
But it has one defect:
It needs a mechanic.

General, man is very useful.
He can fly and he can kill.
But he has one defect:
He can think.”

Finally generals no longer need a (human) driver or mechanic!!
The AI driver can think, yes; but we/generals can decide and control HOW it will think!

Bertolt Brecht (1898-1956)
“Engineeering Moral Agents”: Dagstuhl Seminar

“Imbuing robots and autonomous systems with ethical norms and values is an increasingly urgent challenge, given rapid developments in, for example, driverless cars, unmanned air vehicles (drones), and care assistant robots.”

Two immediate problems.

➢ the **formalisation of ethics** in a format that lends itself to machine implementation;
➢ the actual **implementation of moral reasoning and conduct in autonomous systems**
The mass media’ **PROBLEMS** are mainly:

- Privacy
- Security (on WEB, ... on access ..
- Anthropomorphism
- War and Artificial soldierds/arms
- Ethics *inside* Artificial creatures and algoritms

For me

**not less serious problems**....
Is our Intelligent Technology research ONLY BUSINESS ORIENTED just because it needs money???
Meeting of the minds for machine intelligence

Industry leaders, computer scientists and students, and venture capitalists gather to discuss how smarter computers are remaking our world.

Once a machine is educated, it can help experts make better decisions

... savvy machines can help us evaluate (social) policies. Etc...

Are ONLY THESE THE RIGHT SUBJECTS/MINDS TO INVOLVE ??? for discussing about ethical and political and social consequences of machine intelligence and hybrid society???

What about other subjects to be involved like: moral and political philosophers, trade unions, social democratic movements (like women movement, like “occupy Wall Street”..), poor countries, etc.
Meeting of the minds for machine intelligence

Industry leaders, computer scientists and students, and venture capitalists gather to discuss how smarter computers are remaking our world.

Once a machine is educated, it can help experts make better decisions

... savvy machines can help us evaluate (social) policies. Etc...

Why alliance only between academy, scientists, and capitalists and business men??

Is this so OBVIOUS and UNDISPUTABLE in current culture to become INVISIBLE?
Meeting of *the minds for machine intelligence*

Industry leaders, computer scientists and students, and venture capitalists gather to discuss *how smarter computers are remaking our world*.

Once a machine is educated, it can help experts make better decisions...

... savvy machines can help us evaluate (social) policies. Etc...

“Better” for whom?

It is not a “technical” problem, but a political problem. “Better” for poor and powerless people/countries of for dominating classes, lobbies, powers, countries?
Carnegie Mellon University plans to announce on Wednesday that it will create a research center that focuses on the ethics of artificial intelligence. The ethics center, called the K&L Gates Endowment for Ethics and Computational Technologies, is being established at a time of growing international concern about the impact of A.I. technologies.

That has already led to an array of academic, governmental and private efforts to explore a technology that until recently was largely the stuff of science fiction. ... Peter J. Kalis, chairman of the law firm, said the potential impact of A.I. technology on the economy and culture made it essential that as a society we make thoughtful, ethical choices about how the software and machines are used.
Just an other example of this **obvious alliance**:

The ex-CEO (Chief Executive Officer) of Google and current **President of Google CdA**, Eric Schmidt, has been nominated by Ash Carter, USA Defense Secretary, **chairman del DoD Innovation Advisory Board of Pentagon**.

The aim of DoD Innovation organization is the **full exploitation of Sylicon Valley **"best practices"** for military purposes**

**AGAIN:**

Why an **alliance only** between academy, scientists, and capitalists and business men, (and war powers)??

Is this so OBVIOUS and UNDISPUTABLE in current culture?

**Is our Intelligent Technology research **ONLY BUSINESS ORIENTED** just because it needs money??**
Hidden Interests
Security, Privacy, War, Ethics, .. are for sure very relevant issues, we have to reflect on,

    BUT not the most or the only relevant ones from the moral and political point of view.

Hidden interests, manipulation of us (users and programmers), exploitation, ... emptying democracy, etc. are NOT less important.

We have to be conscious

    not just manipulated, unaware although genial servants of those forces and interests.

Democracy is not a formal and misinformed voting ritual.

    WE have to foster a real “intelligence” (understanding) and EMPOWERMENT of people in/on the hybrid societies evolution.
“MANIPULATION” Theory:

X influences Y’s behavior by communication or action, by changing Y’s goals, by changing his/her beliefs and feelings, but in a hidden way.

That is, without communicating his INTENTION of INFLUENCING Y, of CHANGING Y’s mind.
Not only improved and collective INTELLIGENCE but improved and collective AWARENESS, which is a crucial form of “intelligence”, of understanding what we are doing and WHY we are doing that; who is “nudging” us.
Not only improved and collective INTELLIGENCE but improved and collective AWARENESS,

Intelligent Agent and algorithms have to help us to understand not only our Goals and how to decide (by revealing and correcting our rational & affective BIASES) but help us to understand our “finalities”/“aims”, which go much beyond our mental Goals.
HELP in **RATIONAL** DECISION MAKING, but...

The real problem is not that “our” decision be fully efficient and *rational* (not misinformed or biased), but:

in favor of whom? With **AWARENESS** of **“interests”** not only of goals: conscious or declared or *ascribed* to us!!!
“FUNCTIONS” Theory

“External goals” impinging on us and on our minds and conducts.

NON INTENDED “purposes” of our behavior.

We follow them without understanding them, or deciding about them, although they not only “emerge” from our collective behaviors, but feedback and “immerge” in our minds,

(“IMMERGENCE” theory; homage to Rosaria Conte)
Also the Goals of our Agents and Robots (or just their rules and procedures) serve to *FUNCTIONS*: external, not chosen and understood GOALS.

Are they explicit, transparent at least for us?

They favor some *interest*.

*Is this intended or not-intended by us?*

To which *VALUES* do they respond? Perhaps do not shared by us but at least clear, transparent! Or obscure?
"INTERESTS" Theory
What is better for me and my goals but...
... I do not understand or intentionally pursue them.

Tutelary Role Theory:
X takes care of my "interests", of my good, even in conflict with me, with my current goals; X helps me or pushes me or obliges me!
In a lot of circumstances Agents (or Agent sellers?) will decide *for us* (delegated or not by us), in a **TUTELARY ROLE**

Even if they would just use hints and prompts or just a little push (the celebrated liberal “nudges”) *like in marketing*, by exploiting how the stimulus is presented, the elicited impulses or associations...

1. does really this make us more “free” than an explicit recommendation or an imperative?
Moreover:

2. for whose benefit/advantage?

Who is judging what is better for me, or for us?

Is this really “in our interest” or primarily in the INTEREST of financial and informational dominant powers?

Or (in many countries) of the political regime?
This holds also for more explicit influencing devices like

**RECCOMENDER SYSTEMS**

which will know us better then us.

Will they give us recommendations and suggestions “in our INTEREST”, in a TUTELARY attitude,

or will they follow market criteria

just a more effective, personalized advertising?

On the side of the “user”! Not just of the “seller” (of our data or of some good)
Moreover: “TUTELARY” doesn’t mean “protecting me”

only caring of our “individual” “personal” interests, but also helping us to understand and take care of:

- of **Common interests** and possible **collective subjects** and communities and pressures;
- of hidden **conflicts of interests**;
- of the “commons”, of public goods and their relevance and respect (environment, energy, water, ...)

.
“AUGMENTED INTELLIGENCE”
also means

AUGMENTED **SOCIAL AWARENESS**:

HOW does it work the **“INVISIBLE HAND”** (the god of liberalism)
which organizes the emergent and “spontaneous” social “order”.
The “Mouth of Truth” Algorithm
Clearly we are developing algorithms for ascertaining the “truth” in that mess of data, of assertions, hoaxes, and news, which will be diffused and accessible through the WEB.

An Algorithm for deciding about reliable sources, credible information, what is “true” among so many different claims and data.

There is no alternative on that. However:
On which base such algorithm will “ascertain what is true”? Only on the basis of reliable and convergent sources? Of their number and net? On direct or indirect access to the “fact”? 

Also on the basis of the “values” and on the sharing and acceptability of the values of the source?

Even for ‘official’ science: is it always capturing or saying the truth?

And there will be **dogmatic truths** and **undisputable authorities**, like in any culture?
The people will believe what the media tells them they believe.
- George Orwell
• And **which culture and values** will be assumed as the “right” ones?

How will we allowed to distinguish between a *conflict of values or of interests* from a mere conflict between more or less credible data, more or less grounded, direct, controlled, reliable, ..?
‘PRESENSES’
in our
MIXED REALITY and SOCIETY
The autonomous and proactive intelligent entities that Ag-MAS community created will become ‘presences’ and ‘roles’ in our hybrid society (human and artificial agent) and mixed and augmented reality (combined virtual and ‘real’, ‘natural’ and automatic/prosthetic world).

Now the problem will be:

are we able to manage these autonomous and too informed and intelligent agents?

It is not only a matter of level of delegation (open vs. restricted and specified) or of the possibility of initiative in coop and “over-helping” and even “critical-help” and even functional norms violations; or a matter of adjustable autonomy and negotiation of the degree of power based on trust and self-trust.
It is a matter of:

➢ A) Which **roles** will those material or immaterial, visible and invisible “entities” play in our life and environment?
A) Which **roles**

Will they be **our Guardian angel** with a ‘tutelary’ role?

By helping, protecting and empowering us
A) Which *roles*

Or – less religiously – our Jiminy Cricket *(The Talking Cricket)*
with its recommendations?
A) Which **roles** or our **supervisor** in the **ICT-Panopticon** we live in?

**The Prisoner and the Free**

- can’t see the watcher
- knows he’s being watched
- rounded back
- all alone within a crowd of people
A) Which roles or our tempting Spirit
A) Which **roles** or our **tempting Devil**:

for the benefit of some marketing policy or monopoly, or the influencing and manipulating manager for hidden political or economic powers?
MIXED REALITY,
MIXED BODY & MIND

Will we "incorporate", feel them as part of "us", our "mental prosthesis"?

Will we listen to that moral or rational "voice" as our own mental or consciousness voice our (expanded) SuperEgo.
MIXED REALITY,
MIXED BODY & MIND

Or will our Super Ego be “externalized”?
Not “me”.
Will we listen to "her" as to the voice of our mother, our teacher?
Or will we become “voice hearers”??

Or will it be a boring "Talking Cricket" trying to correct us: wooden marionettes?
But with the advantage that we can to turn off that voice!!
MIXED REALITY, MIXED BODY & MIND

Both solutions will be probably there:

> The “social” one: Externalized voices and Agents

&

> The “reflexively social” one: an augmented internalized Self and Consciousness

Will I prefer to maintaining my judgment and to discuss with some friend or advisor, and listen to him/her or do not pay attention to?? And has s/he to have my own Values and character??

OR

I change and acquire a new Self with new introjected conflicts, Values and style??
PRESENCES

It is a matter of:

➢ B) Which interest they will care of?

Of the interest of the seller or of us as consumers?
Of the interests of the dominant groups or of the submitted or discriminated people?

Which political and moral views will orient them: not our “car driver” but the “society drivers” !!
and our life-navigator.
They will **decide "for us"**, but... AMBIGUOUS: “instead of” us or also “for our good”????

_____________

Social Robots and Intelligent Agents **will NOT govern in their own interest** (science fiction!) but... in the interest of whom????

**EMPOWERING whom????**

And will we be able to monitor and understand that?

And to **make that “transparent” to people?**
VERY RELEVANT to support

“AGREEMENTS” technologies

Intelligent systems for:

- Rational and effective argumentation, negotiation, compromises, fair business practices,..
- Norms, commitments, and value management,..
- Consumer protection, transparency, trust, etc.

However, AGREEMENT Technologies implicitly have TWO SIDES; LET'S look more directly at the other important side:

DISAGREEMENT TECHNOLOGIES
There is a too strong ideology and rhetoric about society as cooperation, collaboration, common intent, collective advantages,.....

how to reach convenient agreements and equilibrium, etc.

Moreover, the web is (non accidentally) favoring a deviating political feeling: “we” against “them” (governors, political caste, centralized powers).

This perception of “we” is completely misleading: there is no a “we” with common values and goals and interests, which has to be unified against the political power as such (in case against the real power (financial power) that has usurped the political power).
A) Population is composed of different classes, genders, generations, .. and cultures with very different and conflicting values and interests; this is the real conflict (not “we” and “them”), and political activity and forces were supposed precisely to represent and protect those different social interests, and not just the “common” interest.
A)

Some conflict of interest or of value can be solved and reconciled in a **common interest**, but a large part of political/government decision is **not for a common advantage** (except reducing civil war), for a fair distribution; it is for **the prevalence or advancement of the interests of a given group** (class, lobby, gender, view, ...) by reducing the powers of the others.
• Making conflicts to emerge and become aware, making express disagreement, making transparent which interests are hidden and prevailing, ... should be (in democracy) one of the main tasks of intelligent social technologies.

Sometimes in order to facilitate a balanced conflict resolution, agreement; sometimes just for supporting the opposition and expressed disagreement that is the ground of democracy.

• Using WEB technologies for organizing “movements” it is OK; but not so good without promoting critical consciousness, understanding of real social interests and hidden powers and conflicts.

• To provide new environment, contexts, and instruments for promoting motivated/grounded CONFLICTS, for making them well grounded and effective, for solving them not just by persuasion but by some achievement.
The Need for Conflicts

Conflicts: the presupposition of Democracy

No conflicts no democracy

Democracy is not only a "response" to Cs and for moderating them; it would be a way of encouraging, growing (and solving). Conflicts are not only to be governed, reduced, reconciled: they should even be promoted and this is in fact the role/function of specific forces and organizations, like trade-unions, parties, group of interests, associations, movements, etc. Crucial stakeholders of democracy, but also definitely responsible of the typical social, cultural, economic "progress" of western countries in the last centuries and now of the rest of the world.

Of course conflicts might be dangerous conducting us to fighting, violence, war, .. So it is true that societies and groups need "rules" for governing them, to avoid degeneration. Centralized state was one of these solutions: the state monopolizes violence; private or group violence is forbidden.
Conflicts: the presupposition of Democracy

Conflicts are not just conflicts of views or opinions, or due to different conceptions, information, reasoning.

There are conflicts of “objective interests”: if you realize your goal I cannot realize my goal or loose something I have.

So the problem is conflicts between interests of group or classes, or conflicts between "private" interests vs. common interests, the "commons" and public goods.

Social conflicts in fact do not have a "verbal/cognitive" or a "technical" solution, just based on data and technical principles; they have a "political" solution; it is a matter of "power" and of prevailing interests and compromises (equilibrium, partitions/shares).
Conflicts with their disagreements and agreements are thus the motor and principle of Democracy and of its possible effectiveness in changing society in favor of the submitted subjects, disadvantaged classes and groups, etc.

Viva conflicts!
“LIFE navigator”

B) ICT and cognitive technologies are strongly submitted to the private interest of marketing; they are used for recognize your profile and interests but NOT for EMPOWERING YOU, but in order to propose/induce you to “buy” something (goods, ideas, ..) They are monitoring and analyzing you in order to manipulate you and influence your choices.

We need anti-manipulation AI technologies:

I would like to have not so much a personal virtual or robotic psychotherapist or physiotherapist;
We need **anti-manipulation** AI technologies:

I would like much more a “**life navigator**” in my main “social role” (ex. consumer!), but not a navigator saying “turn right, turn left”, “buy that; do not buy this”...

But **a tutor, a trainer**, inducing me to **understand** and to reflect about why I’m oriented in that direction, I’m choosing that product; worrying if I have the right information, or I have wrong beliefs, etc.

**Making me conscious of who and how** is persuading or just unconsciously manipulating **me**; and so on.

Mixed reality not only with some '**tempting devil**’ but with some virtual ‘**guardian angel**’ (work with Alessandro Ricci and Luca Tummmolini)
C) We need environments and Agents for learning and developing a “critical thinking” attitude; to manage our cognitive and motivational biases; etc.

To support us in argumentation and discussion, and in understanding the tricky arguments of the others.
To resist to the prevalence of “audience” against “quality”, of self-marketing and indexes against originality and quality; etc.

And so on, about propaganda, Academy, gender models, fanaticism, superstition, urban legends, ...

We have impressive possibilities with new intelligent and interacting technology, big data, etc.
They shouldn’t be just used for selling and for dominating.
Concluding Remarks
Self-Organization = Out of Minds

Society works thanks our *partial* intelligence.

We cooperate and jointly act not just in spite of but *thanks to the fact that we do not* (fully) understand and intend what we are *jointly doing.*

And even if we disagree.

**Society is governed by the “Invisible hand”**

Obviously there are also very dysfunctional and undesirable results of our stupidity/blindness.

For example, *since we do not understand or forget, and thus do not learn from the worse tragedies of our History,* that's why we repeat them as farces.
“Alienation”:

We are dominated by our own ‘delegated’ (emergent) (social and intellectual) powers, we are not aware of;

we do not realize and we do not decide/intend to create such collective or such institutional powers impinging on us.

Society is governed by the “Invisible hand”
Can We Overcome our *Alienation*?

Will the **Leviathan** become a giant *connected and informed* community of *agents*, managing their Collective Power?

1. I’m skeptical about that (also for cognitive reasons)

2. I worry about possible *net-Demagogy*
Can We Overcome our **Alienation**?

Could we, by exploiting

- collective, distributed, hybrid INTELLIGENCE
- and
- BIG DATA
- and
- run-time feedbacks and information from local stakeholders and intelligent sensors
- and
- Computational LEARNING and PREDICTING
- and
- Computer (Agent-based) Social SIMULATION, and VIRTUAL REALITY and SERIOUS GAMES, etc.

could we

**MAKE VISIBLE the INVISIBLE HAND?**

and (partially) GOVERN IT???
In the Digital Society

Artificial Intelligence may either exploit or overcoming our Natural Stupidity
Democracy

Mark Twain is brilliantly right
"If voting made any difference they wouldn't let us do it."

But… the problem is much harder; it is not just a complot, is that we vote in a self-defeating way, and, in general, our collective stupidity.

Might political "education" and education to "commons" & Digital society and participatory democracy be enough, and solve this “cognitive” and social problem?

They will help. But given the immediate local perception of the conflicting interests and competition and the blindness to common interests among different countries and poor classes and ethnic groups, and affiliation and identity feelings, conformism, and in-group vs. out-group psych, … I have some doubt.

In a couple of centuries they will see.
PESSIMISM of the INTELLECT but OPTIMISM of the WILL
Pessimism of a real EXPERT of POWER:

Napoleon Bonaparte:

“When a government depends on bankers for money, they and not the heads of government control matters. The hand that gives holds sway over the hand that receives. Money has no homeland and financiers have no patriotism or decency - their only objective is gain”
Agains Thrasymachus pessimism

"Justice is nothing more than whatever is advantageous to the stronger" Thrasymachus

Can we build some Artificial Intelligence and some Hybrid and Mixed Sociality & Reality smurter than us but on our side, for our collective interests and commons,

such that we can eventually face Thrasymachus’ "unavoidable" result?

Competence and ideas are there, HERE!

What about means?
The GLASS of the INVISIBLE

The great **REVOLUTION** of ICT, of digital *monitoring* and *predicting* (by simulation) can give to society (to demos)

a glass were to observe themselves and follow what it is happening.

**A glass reflecting also what is invisible: hidden presences, the future:**

**A GLASS OF the INVISIBLE**
The GLASS of the INVISIBLE

Of “HERE” and “NOW”

Not only of what is “not present” here, and can be virtually present for interaction, can act in this word and vice versa, etc.

”PRESENCES”: in mixed virtual and real world

But also a glass able to show what cannot be seen/understood: the future, predictions, and hidden phenomena and interests (for example, can I see who is now getting my personal data? And for what?)
To See What Is (currently) Invisible:

*Artificially Augmented Awareness*

The real revolution of AI

Including itself! It uses
SORRY for such a PESSIMISTIC TALK

but

I wish you get the message of

the Optimism the WILL
Thank you for your attention!

And sorry for my “English”
I like to thank our research group in Cognitive Science at ISTC: the ‘GOAL group’

Rino Falcone
(Emiliano Lorini)
Maria Miceli
Fabio Paglieri
Giovanni Pezzulo
Luca Tummolini
.......  

And IN MEMORY of Rosaria Conte (Social Simulation LABSS Group)

http://www.istc.cnr.it/group/goal
Why creation, design, and diffusion of ICT Technologies should just service capital and not workers and people?

How could we create or use technology in order to defend subjects, poor, marginalized, exploited people?

It is enough that capital puts money in research and university for being unaware of its domination and for accepting to be a-critical instruments of its logics.