

SERIOUS VALUES CANNOT BE B**S*****

Quantifying AI Transparency, and UN Sustainability:

AIM North Symposium 2019, November 7 2019 , (15 minutes talk)
<https://www.aim2north.com/Program-2019>

By
TOM GILB

tom@Gilb.com
www.Gilb.com
+47 92066705
Copyright Tom Gilb 2019

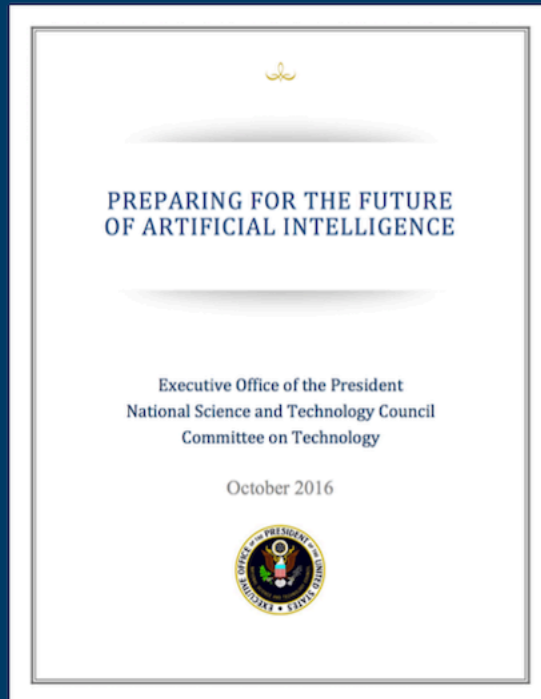


These slides will eventually be placed here:
http://concepts.gilb.com/tiki-list_file_gallery.php?galleryId=14

‘Principles’ = —> ‘Qualities’ ! *
With the *usual lack of quantified definition*

Oct. 2016 : “Preparing for the future of AI” from USA

IBM



Proposal of Discussion toward Formulation of AI R&D Guideline Distributed material

Referring OECD guidelines governing privacy, security, and so on, it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked ("AI R&D Guideline") as framework taken into account of in R&D of AI to be networked.

Proposed Principles in "AI R&D Guideline"

1. Principle of Transparency

Ensuring the abilities to explain and verify the behaviors of the AI network system

2. Principle of User Assistance

Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices

3. Principle of Controllability

Ensuring controllability of the AI network system by humans

4. Principle of Security

Ensuring the robustness and dependability of the AI network system

5. Principle of Safety

Giving consideration so that the AI network system will not cause danger to the lives/bodies of users and third parties

6. Principle of Privacy

Giving consideration so that the AI network system will not infringe the privacy of users and third parties

7. Principle of Ethics

Respecting human dignity and individuals' autonomy in conducting research and development of AI to be networked

8. Principle of Accountability

Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked

AI Accountability Defined

Fuzziness -> Quantified. Structured, Enriched

I did this for all XAI Qualities

needs&means from RSA Technology

AI Qualities... Project Canvas Tables Diagram More... Create Show Sidebar tomgilb

AI Qualities Demo 2019 / Canvas | List | Diagram / AI Accountability

Name Tag

AI Accountability

Level: Stakeholder, Type: Value, Labels: - Edit

Is Part Of: CRITICAL AI QUALITIES

Simple Graphic of the Required Improvement

Long Term Overall AI Objective.Wish [Scenarios = All, AI Suppliers = All, Accountability Strength = All, AI Consumers = All] @ 09 Jan 2040 : 90 ±8% <- tsg

The 'well-intended' statement

Ambition Level: "Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked"

Stakeholders: ACADEMIC INSTITUTIONS, AI Users. there are more stakeholders and more-detailed types, but I just wanted to mention the ones in the

Scale: % of [Scenarios] where [AI Suppliers] have a [Accountability Strength] for an Accountability Type] with relation to [AI Consumers]

Defining qualities quantitatively

Wild Guess At Present Overall AI Status Level.Status: Level: 1 extremely high, little data and experience [Scenarios = All, AI Suppliers = All, Accountability Strength = All, AI Consumers = ...]

Long Term Overall AI Objective.Wish: Level: 90 ±8% [Scenarios = All, AI Suppliers = All, Accountability Strength = All, AI Consumers = All] When 09 Jan 2040

Goal level

AI 'Accountability' scale Detailed [Scale Parameters] Why is this Useful?

I did this for all XAI Qualities

AI Accountability Defined.Scale:

1



+ Add



% of [AI Scenarios] where [AI Suppliers] have a [Accountability Strength] for an [Accountability Type] with relation to [AI Consumers].

Templates

+e to add qualifiers ? / terms ? / specs ? , formatting

AI Consumers: defined as:

Citizens, Medical, Transport, Government, Telecoms, Finance, Industry, Commerce, Distribution, Academic, Entertainment, Autonomous Vehicles, IOT Consumers, Military, Police, Senior Homes, Private Homes, Apartments, Offices...

AI Scenarios: defined as:

Freeware, Paid Services, **[Life Critical Decisions]**, Health Decisions, Large Scale Decisions, Long Term Decisions, Entity Survival Decisions, Political Choices, Games...

AI Suppliers: defined as:

App Developers, Data suppliers, Academic Research, Startups, Consultants, IOT suppliers...

Accountability Strength: defined as:

Legally Compliant, Contractually Compliant, Standards Compliant, International Guidelines...

Accountability Type: defined as:

Board Responsibility, Legal Responsibility, Ethical Responsibility, Financial Responsibility, Risk Responsibility , ...

[Scale Parameters]
Help us decompose complex systems
And then to prioritise critical parts of them
For early value delivery

Quantitative Evaluation of AI Technology Qualities using 'Impact Estimation Tables'

EVALUATE AN AI TECHNOLOGY'S QUALITIES

From Level: Stakeholder To Level: Product

Settings... Add Sort Duplicate... Undo... =: ABSOLUTE Help me!

	💡 Ideal XAI Technol...	💡 Notional Technique	💡 UC Berkeley Deep ...	💡 UCLA Pattern Theory+	💡 Charles
Requirements					Show Sideba
AI Accountability Status: 1 → Wish: 90 % of [AI S...Δ%:	100%	56%	????	????	????
AI Controllability Status: 50 → Wish: 98 % of [AI S...Δ%:	100%	88%	????	????	????
AI Ethicality Status: 5 → Wish: 90 % #Respect...Δ%:	100%	-4%			
AI Privacy Status: 1 → Tolerable: 0.5 % of [Priv...Δ%:	100%	80%			
AI Robustness Status: 70 → Wish: 99.9 % #Intende...Δ%:	100%	33%			
AI Safety Status: 100 → Wish: 10 Number of Δ%:	100%	47%	????	????	????
AI Transparency Status: 10 → Wish: 99.998 % of selec...Δ%:	100%	47%	????	????	????
AI Usability Status: 1 → Wish: 0.1 Speed in M...Δ%:	100%	33%	????	????	????
AI Learning Performance Status: 0 → Wish: 50 % [Learnin...Δ%:	100%	84%	????	????	????

This is a quick mockup without real data:
to show the potential for evaluating AI
Techniques
and their effects on
AI Values (or 'Qualities')

AI 'Technique Evaluation'

with objective evidence, sources uncertainty

EVALUATE AN AI TECHNOLOGY'S QUALITIES

From Level: Stakeholder To Level: Product

Settings... Add Sort Duplicate... Undo... =: ABSOLUTE Help me!

Requirements		Ideal XAI Technol...	Notional Technique	UC Berkeley
AI Accountability Status: 1 → Wish: 90 % of [AI S... % of [AI Scenarios] where [AI Supplie...	=: Δ%:	100%	56%	????
AI Controllability Status: 50 → Wish: 98 % of [AI S... % of [AI Situations] and [AI Accident...	=: Δ%:	100%	88%	????
AI Ethicality Status: 5 → Wish: 90 % #Respect... % Respect Violations of a [Respect Vi...	=: Δ%:	100%	-4%	????
AI Privacy Status: 1 → Tolerable: 0.5 % of [Priv... % of [Privacy Breach Types] annually ...	=: Δ%:	100%	80%	????
AI Robustness Status: 70 → Wish: 99.9 % #Intende... % Intended Uptime for [AI Consumers] ...	=: Δ%:	100%	33%	????
AI Safety Status: 100 → Wish: 10 Number of ... Number of annual AI Damage occurrenc...	=: Δ%:	100%	47%	????
AI Transparency Status: 10 → Wish: 99.998 % of selec... % of selected [AI Cases] where we can...	=: Δ%:	100%	47%	????
AI Usability Status: 1 → Wish: 0.1 Speed in M... Speed in Minutes for an [AI User Type...	=: Δ%:	100%	33%	????
AI Learning Performance Status: 0 → Wish: 50 % [Learnin... % [Learning Performance] improvement ...	=: Δ%:	100%	84%	????
Sum Of Values:	Σ%:	900 %	464 %	0
Operational Cost Budget: 0 → Status: 100 % of a def... % of a defined Budget	=: Δ%:	99%	99%	
Sum Of Development Resources:	Σ%:	99 %	99 %	0
Value To Cost:		9.10	4.70	

Hide Sidebar

+Specification

-Impact Cell

Value: AI Accountability
Solution: Notional Technique
Scale: % of [AI Scenarios] where [AI Suppliers] have a [Accountability Strength] for an [Accountability Type] with relation to [AI Consumers].

Tag.Value Impact:

Estimate: 50 ± 10
Actual: scale value ± 0
Credibility:
Evidence: MEASURED IN PHD STUDY, 3 SITES
Source: by tomgilb - Jan 10th 2019, 04:16
UCLA URL
Templates

+Rules & Processes

Main point =
When AI qualities are quantified
And the Wished level is specified
Then we can evaluate various AI technologies
And find the most cost-effective overall

↔ = Scale of measure for all the other AI Values

My (TSG) Observations and Conclusions

- The most fundamental obstacle to AI Standards progress is WE NEED TO **QUANTIFY AND STRUCTURE-RICHLY** all **CRITICAL STAKEHOLDER VALUES (ALL > 8)**
- Next: we need much-deeper, more-comprehensive identification, and consensus, and detailed knowledge (their values and constraints) about AI/ AI STANDARDS **STAKEHOLDERS**

Proposal of Discussion toward Formulation of AI R&D Guideline

Referring OECD guidelines governing privacy, security, and so on, it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked ("AI R&D Guideline") as framework taken into account of in R&D of AI to be networked.

Proposed Principles in "AI R&D Guideline"

1. Principle of Transparency

Ensuring the abilities to explain and verify the behaviors of the AI network system

2. Principle of User Assistance

Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices

3. Principle of Controllability

Ensuring controllability of the AI network system by humans

4. Principle of Security

Ensuring the robustness and dependability of the AI network system

5. Principle of Safety

Giving consideration so that the AI network system will not cause danger to the lives/bodies of users and third parties

6. Principle of Privacy

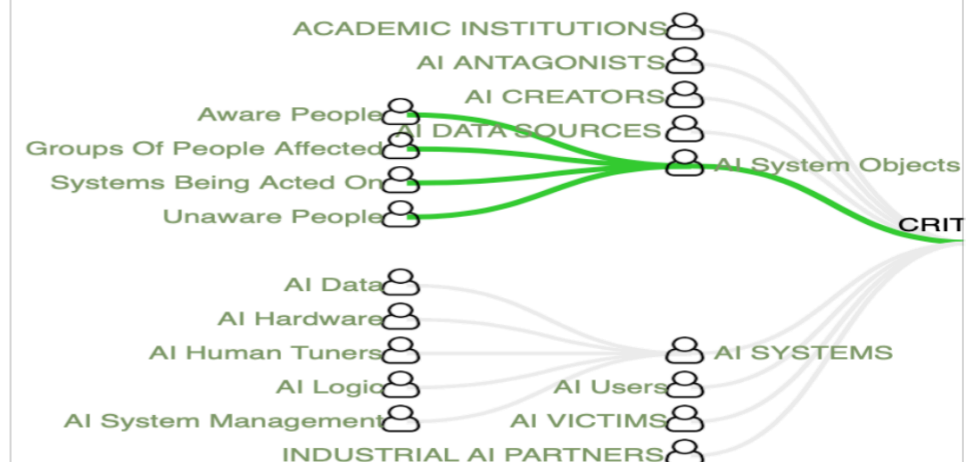
Giving consideration so that the AI network system will not infringe the privacy of users and third parties

7. Principle of Ethics

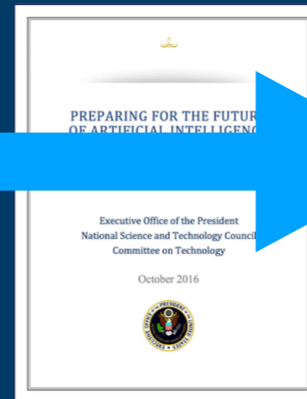
Respecting human dignity and individuals' autonomy in conducting research and development of AI to be networked

8. Principle of Accountability

Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked



Details of Quantifying all the XAI Objectives



Proposal of Discussion toward Formulation of AI R&D Guideline

Referring OECD guidelines governing privacy, security, and so on, **it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked (“AI R&D Guideline”) as framework taken into account of in R&D of AI to be networked.**

Proposed Principles in “AI R&D Guideline”

- 1. Principle of Transparency**
Ensuring the abilities to explain and verify the behaviors of the AI network system
- 2. Principle of User Assistance**
Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices
- 3. Principle of Controllability**
Ensuring controllability of the AI network system by humans
- 4. Principle of Security**
Ensuring the robustness and dependability of the AI network system
- 5. Principle of Safety**
Giving consideration so that the AI network system will not cause danger to the lives/bodies of users and third parties
- 6. Principle of Privacy**
Giving consideration so that the AI network system will not infringe the privacy of users and third parties
- 7. Principle of Ethics**
Respecting human dignity and individuals' autonomy in conducting research and development of AI to be networked
- 8. Principle of Accountability**
Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked

© 2017 International Business Machines Corporation

17

XAI Explaining AI
Lecture Slides

<http://concepts.gilb.com/dl958>

A Serious ‘Multi-dimensional Metrics Attack’ on
Poor AI ‘Academic and Standards’ Thinking & Planning.

An analysis of published Principles for Managing and
Standardizing AI,

where about 10 AI Qualities like ‘Safety’ and
‘Transparency’ are shown to be quantifiable.
This is prelude to rational thinking about the entire subject.

GilbFest Talk June 25 2019
Presented to the IBM AI Standards Committee Members
(who said they were suitably impressed)



Do politicians like BS?

If you bullst an AI system, could you fool it into making bad decisions?**

- **Sustainability Planning**
- <https://tinyurl.com/UNGoalsGilb>
- **Free book, with details of how to analyze, clarify.**



Transparency

We are acutely aware of the privileged position we occupy as a monopoly provider of an essential service. We are a top performing UK DNO group and accept that this high performance brings with it increased scrutiny from our customers, other stakeholders and the industry at large. We are committed to being open about our operations and, through documents such as this Annual Review and our RIIO-ED1 commitments report, we seek to make it easy for all our stakeholders to understand how we are performing and, most importantly, how we are performing for our customers.

We believe that it makes business sense for our employees to be incentivised to deliver on our vision. The main way employees are incentivised is through a bonus scheme (scheme). In line with our vision, our employees are rewarded for delivering on our vision, customer service and cost saving.

Since its inception in 2010 the key focus of UK Power Networks has been to deliver what our customers want at the lowest possible cost. We have delivered significant improvements in the reliability of our

On this page you will see how we have performed against the following Sustainable Development Goals:



My associates are using these UN Goals as a framework.
But, do all stakeholders have an identical clear understanding of the goals?

90%
of UK Power Networks employees are covered by the Trade Union Collective Agreements

End poverty in all its forms everywhere.



Nice but not specific or measurable

Nice but not specific or measurable

Goal 1. End poverty in all its forms everywhere

Target 1.5: By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters

- **Indicator 1.5.1:** Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population
 - See Metadata :
- **Indicator 1.5.2:** Direct economic loss attributed to disasters in relation to global gross domestic product (GDP)
 - See Metadata :
- **Indicator 1.5.3:** Number of countries that adopt and implement national disaster risk reduction strategies in line with the Sendai Framework for Disaster Risk Reduction 2015-2030
 - See Metadata :
- **Indicator 1.5.4:** Proportion of local governments that adopt and implement local disaster risk reduction strategies in line with national disaster risk reduction strategies
 - See Metadata :

Attempt to find something measurable
But
IS THIS REALLY THE CORE OF
POVERTY CONCERNS

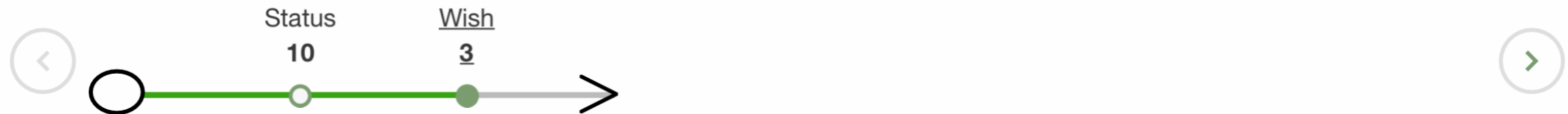
No specific targets set

Simple Goal 1 Draft

1. Poverty (Simple)

Show Sidebar

Level: Product, Type: Value, Labels: - Edit



Wish [Poverty Form = **Financial Income**, Area = {**Central America, South America**}, Poverty Scale = % **Below Minimum Financial Income Per Day in \$**] @ 2030 : **3** \$ <- Tsg

Ambition Level: End poverty in all its forms everywhere

Stakeholders: UN SDG Partners, UN Sustainability Development Goals, United Nations.

Scale: % [Poverty Form] in [Area] as quantified by [Poverty Scale]

Status: 10 \$ [Poverty Form = **Financial Income**, Area = {**Central America, South America**}, Poverty Scale = % **Below Minimum Financial Income Per Day in \$**] Wh.

Wish: 3 \$ [Poverty Form = **Financial Income**, Area = {**Central America, South America**}, Poverty Scale = % **Below Minimum Financial Income Per Day in \$**] When .

- Problem. This spec is too simple; it cannot serve the purposes of modelling critical aspects of complex systems.



Facts and Figures

Goal 1 Targets

Links

1.1 By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day

1.2 By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions

1.3 Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable

1.4 By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance

1.5 By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters

See next slides, for analysis of this

1.A Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and policies to end poverty in all its dimensions

1.B Create sound policy frameworks at the national, regional and international levels, based on pro-poor and gender-sensitive development strategies, to support accelerated investment in poverty eradication actions

Values and Resources

- 1. Poverty (Decomposed)
 - 1.1 Financial Poverty
 - 1.2 Poverty In All Dimensions
 - 1.3 Social Poverty
 - 1.4 Resources. Poverty
 - 1.5 Disaster Protection Po**
 - 1.A National Poverty Reduct...
 - 1.B Investment Frameworks...

Ok Let's take Poverty Target 1.5 as an example

“By 2030, **build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters**”



<- see next slides for analysis and re-organization of this



% #Success Level# in [Building] [Resilience] for [Vulnerable] in [Situations] to [Shocks].

Templates ▾

Building: defined as:

Economic Power, Health Power, Communications Ability, Recovery Speed, Relocation Capability, ...

Resilience: defined as:

Avoiding, Escaping, Resisting, Recovering, ...

Shocks: defined as:

Climate, Economic, Social, [Environmental]

Environmental: defined as:

Earthquake, Flood, Avalanche, Fire

Situations: defined as:

Individual Poverty, Family Poverty, Communal Poverty, National Poverty, Epidemic Hit,

Success Level: defined as:

The attainment of Resilience for the defined circumstances. EG Avoided %.

Vulnerable: defined as:

Poor, Physically Exposed, Weak Health, No Network Fallback, Insufficient Insurance, Insufficient Savings, Employment Problems, .

Defining a structured scale of measure

- The Scale is defined as a set of 'Scale Parameters'
- Each SP is defined as a set of conditions
- Each condition can be defined
- Terms can be defined as needed



tag:SECURITY

Ambition level:

Make Warsaw safe again, by reducing and mitigating any attacks or threats to citizens and visitors

Scale:

% Known Annual

[Attack Types]

having

[Potential Or Real Effects]

on specific

[Stakeholders]

Fake News

Corporate System Down

Attackers

Random Lock Files

Government System Down

Client of Attackers

Denial of Service

National Services Down

Attack Individual Victims

Steal Secrets

Military Systems Down

Corporate Victims

Steal Marketing
Information

...

Government Victims

Pranks or PR Security
Attacks

...

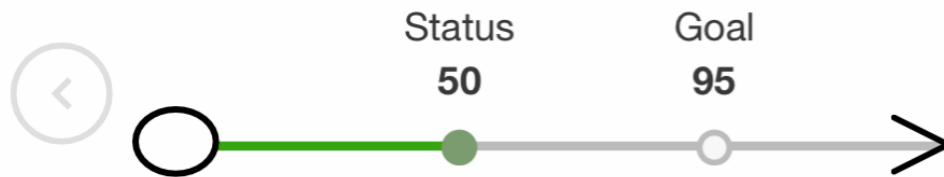
- Visualization of a Scale, with Scale parameter 'dimensions'

1.5 Disaster Protection Poverty

Level: Product, Type: Value, Labels: - Edit

Is Part Of:  [1. Poverty \(Decomposed\)](#)

Show S



Health .Status [Building = **Health Power**, Resilience = **Escaping**, Vulnerable = {**Physically Exposed, Weak Health**}, Situations = **Family Poverty**, Shocks = {**Climate, Environmental**}] @ 2019 : **50** % Escaped



Tag.Ambition Level:



“By 2030, **build** the **resilience** of the **poor** and those in **vulnerable situations** and reduce their **exposure and vulnerability** to **climate-related** *extreme events and other economic, social and environmental* **shocks and disasters**”

- **Now we can apply the scale and some selected Parameters to define the status quo, and the Goal level**

1.5 Disaster Protection Poverty

Show Sidebar

Level: Product, Type: Value, Labels: - Edit

Is Part Of: ➡ [1. Poverty \(Decomposed\)](#)



Ambition Level: “By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-.

Scale: % Success Level in [Building] [Resilience] for [Vulnerable] in [Situations] to [Shocks].

Stakeholders: UN Children’s Fund, UN Global Compact, UN International Strategy For Disaster Reduction.

Health .Status: 50 % Escaped [Building = **Health Power**, Resilience = **Escaping**, Vulnerable = **Weak Health**, Situations = **Family Poverty**, Shocks = {**Climate**, **Env.**]

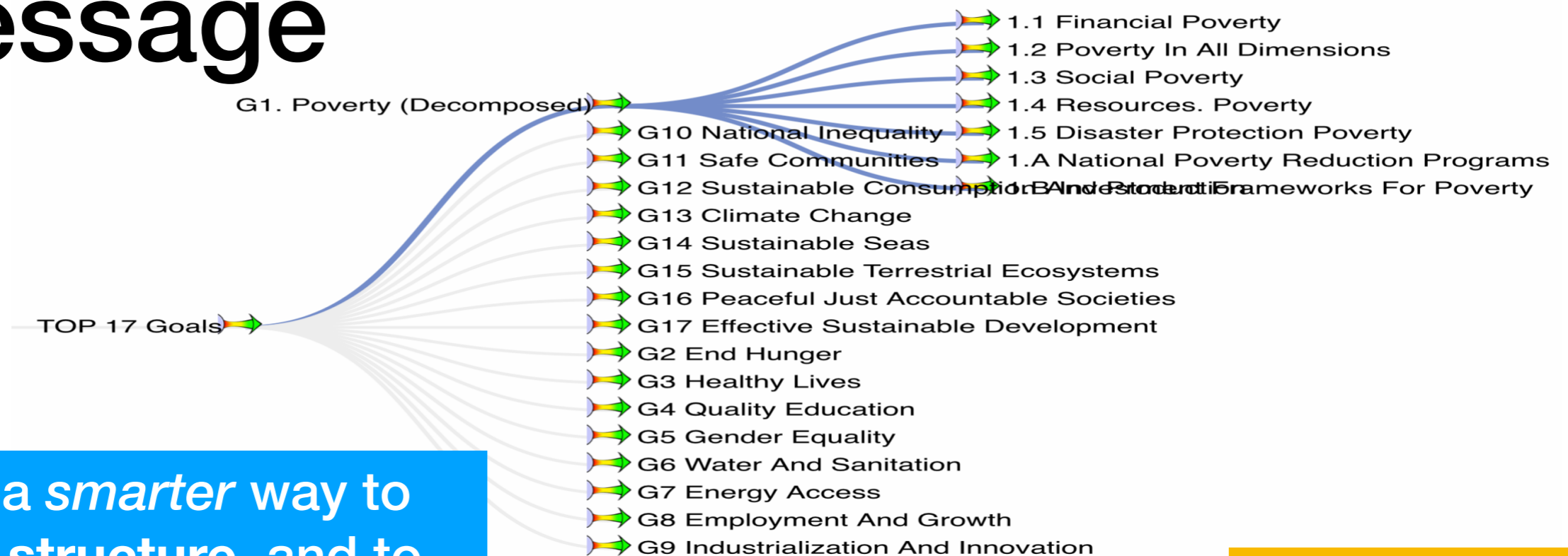
Health .Goal: 95 [Building = **Health Power**, Resilience = **Escaping**, Vulnerable = **Weak Health**, Situations = **Family Poverty**, Shocks = {**Climate**, **Environmental**}] Wh.

Overall Long Term.Status: 5 % [Building = **All**, Resilience = **All**, Vulnerable = **All**, Situations = **All**, Shocks = **All**] When 2019

Overall Long Term.Goal: 98 % $\pm 1\%$ [Building = **All**, Resilience = **All**, Vulnerable = **All**, Situations = **All**, Shocks = **All**] When 2030

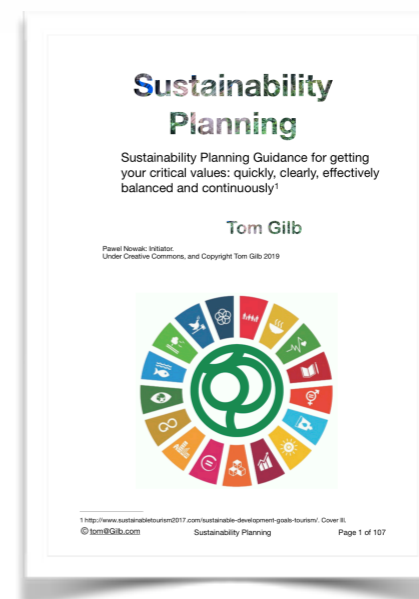
Setting long and short term Goals, re-using the Scale of measure
Using a variety of scale *levels* (Wish, Goal, Stretch, Tolerable, Past, Status), and [Scale parameters].
To articulate: How much ‘poverty’ where, when, for whom, which type, under which conditions.

I am done with my main message



“There is a *smarter* way to clarify, to structure, and to quantify all critical value objectives




If you do *not* clarify, you will for sure *waste your time* trying to get the results you dream of”





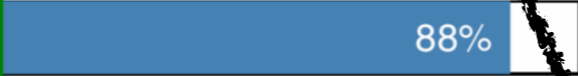



More detail?
“Sustainability
Management”
Digital Book 2019
[https://
tinyurl.com/
UNGoalsGilb](https://tinyurl.com/UNGoalsGilb)
Later see
Gilb.com

Now if I have time I'd like to give you one insight into some advantages of this kind of goal quantification and structuring

- Main idea
- You can now evaluate your strategies
- Quantitatively
- In many dimensions
- With regard to risks
- And costs

your		 <u>T1.5.1 Flood Dama...</u>
Requirements		
 <u>1.5 Disaster Protection Pov...</u> Status: 50 ➔ Goal: 60 % #Succes... % Success Level in [Building] [Resil... [Building = Health Power  2030	Δ : = $\Delta\%$:	<u>5</u> 55 % #Suc... 50 % <div><div>50%</div></div>

Requirements		💡 T1.5.1 Flood Dama...	💡 T1.5.2 Flood Temp...
 1.5 Disaster Protection Pov... Status: 50 → Goal: 60 % #Succes... % Success Level in [Building] [Resil... [Building = Health Power 📅 2030	Δ: =: Δ%:	5 55 % #Suc... 50 % 	5 55 % #Suc... 50 % 
Sum Of Values:	Σ%:	50 %	50 %
 Annual Cost Per Dwelling Status: 5 → Budget: 1 Cost for t... Cost for the strategy as % of Dwellin... <i>No qualifiers</i> 📅 2030	Δ: =: Δ%:	-3.5 1.5 Cost fo... 88 % 	-0.5 4.5 Cost fo... 13 % 
Sum Of Development Resources:	Σ%:	88 %	13 %
Value To Cost:		0.60	3.80

We can now estimate how cost-effective our strategy options will probably be.

1.5 Disaster Protection Pov.

Status: **50** → Goal: **60** % #Succes... =:
 % Success Level in [Building] [Resil... Δ%:
 [Building = **Health Power**
 2030

um Of Values:

Σ%: **50** %

Annual Cost Per Dwelling

Status: **5** → Budget: **1** Cost for t... =:
 Cost for the strategy as % of Dwellin... Δ%:
 No qualifiers
 2030

um Of Development Resources:

Σ%: **88** %

Value To Cost:

0.60

Comments:

Add Comment...

Hide Sidebar

Select Impact Target

Tag.Value Impact:

Estimate:

Δ

 ±

Actual:

Δ

 ±

Credibility:

0
·
9

Have previously used this design / strategy on this project and measured it

Evidence:

This strategy works with total 55% success, as defined in our plans, for flood damage prevention in at least 100 UN nations

Source: by tomgilb - Sep 7th 2019, 14:55

UN Research Paper on Flood Control 2019

And if there is time for one more slide.

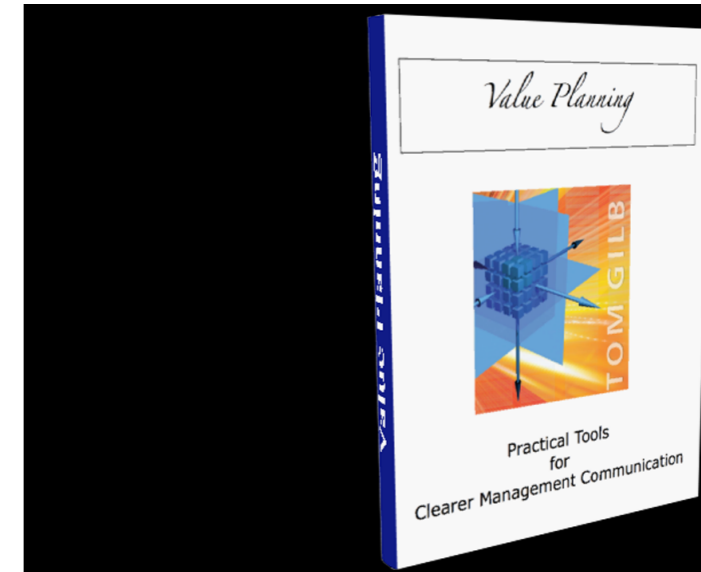
This is how we estimate the credibility of our cost-effectiveness estimates ,
so we know what to trust, and what to prioritise for good results.

More detail?

More detail?
“Sustainability Management”
Digital Book 2019
<https://tinyurl.com/UNGoalsGilb>
Later see
gilb.com

XAI Explaining AI
Lecture Slides
<http://concepts.gilb.com/dl958>

Training ?
www.Gilb.com
Tekna
OSWA
<https://www.meetup.com/Oslo-Software-Architecture/>
(course November 19-20)



Last
Go back one