TRUST
Proposers' Day Briefing
IARPA-BAA-10-03 Overview

Dr. Adam Russell

Intelligence Advanced Research Projects Activity (IARPA)

Office of the Director of National Intelligence (ODNI)

TRUST Disclaimer

- This presentation is provided solely for information and planning processes.

- Does not constitute a formal solicitation for proposals or proposal abstracts.

- Nothing said at Proposers' Day changes the requirement set forth in any BAA

- Any conflict between what is said at Proposers' Day and what is in the BAA will be resolved in favor of the BAA
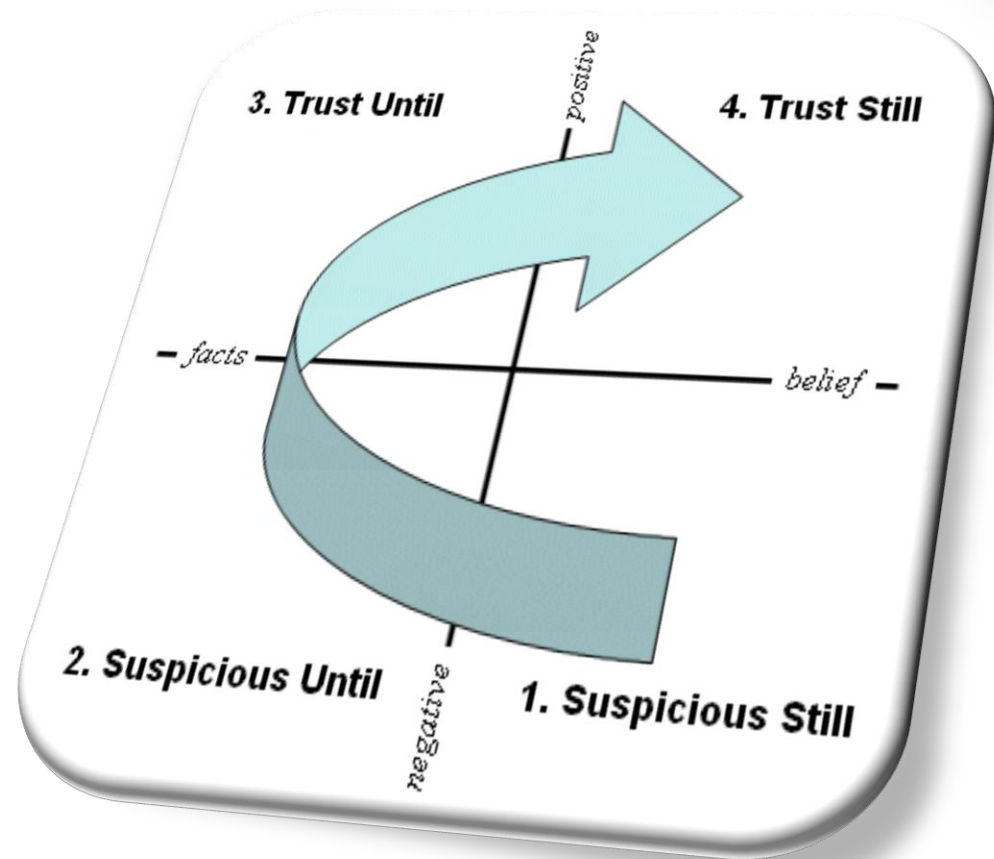
Today

UNCLASSIFIED

Overview: Today's Topics

- Program Overview

- Phase 1

- Phase 2 & 3

- Programmatics

Goals of Proposers' Day

- Listen
- Learn
- Ask questions
- Network
- Save the world

UNCLASSIFIED

dis·ci·pline (ds-pln) n.
- A branch of knowledge or teaching.
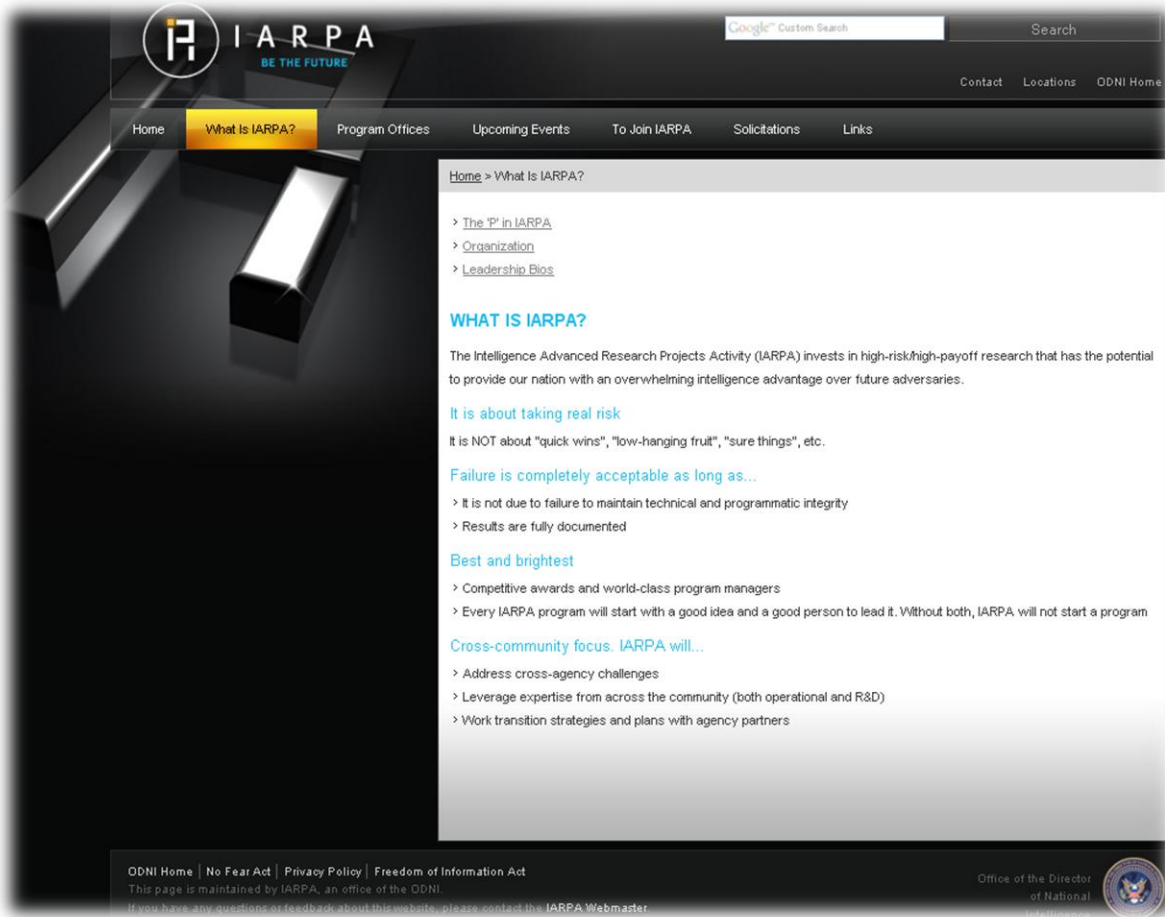
in·ter·dis·ci·pli·nar·y (ntr-ds-pl-nr) adj.
...ng to or involving...

Collaboration is *EXTREMELY* encouraged.

intr.v.
- To work together, especially in a joint intellectual effort.

Program Overview

UNCLASSIFIED

IARPA

Real Risk

Failure is completely acceptable

Best and Brightest

Cross-community Focus

UNCLASSIFIED

Trust and Trustworthiness

- **Trust is a state** associated with a measurable change in the probability that someone will behave in a way that demonstrates their willingness to be vulnerable to an other, in a situation in which their potential benefit is much less than their potential loss if the other person abuses that vulnerability.

- **Trustworthiness is a collection of qualities** of a person that leads them to be considered deserving of trust from others in one or more environments, under different conditions, and to different degrees.

Current limitations:
Why we *don't* trust "trust"

Category	number of studies	accuracy rate
Behavioral (verbal & non-verbal)		
Laypersons (judging friends, romantic partners, adults, children)	N=101	55%
Professionals (judging adults, children)	N=38	56%
Speech Analysis		
Criteria-Based Content Analysis (lab & field)	N=27	70%
Reality Monitoring (lab	N=10	69%
Criteria-Based Content Analysis + Reality Monitoring (lab only)	N=10	74%
Physiological Analysis		
Comparison Question Test polygraph (lab & field)	N=13	75%
Guilty Knowledge Test polygraph (lab & field)	N=8	85%
Guilty Knowledge Test + EEG/P300 (lab only)	N=14	85%
fMRI (lab only; various experimental paradigms)	N=??	78-93%

▶ Highly sensitive to countermeasures

▶ Lack of prospective studies: it's always easier to work backwards, but that rarely moves you forward, especially when ground truth is selective

▶ Detecting deception often provides poor behavioral indicators that are degenerate

▶ Many tools are impractical outside of ideal, sanitized, voluntary conditions i.e. fMRI

▶ Few *validated* or ecologically-relevant model(s) or protocols, i.e. The Guilty Knowledge Test, in which high levels of accuracy are reported but the protocol itself reflects extremely uncommon conditions

Focus on target/deception may be limiting potential capabilities to use other approaches to detect, develop, and sustain trustworthiness

Current limitations:
Why we *can't* (yet) trust "trust"

Psychometrics	Measure	Outcome
Trust antecedents	Meta-analysis of 132 studies on Ability, Benevolence, Integrity	Each antecedent can predict trust and trust outcomes, but only to a very small degree (0.2-.03)
"Quality of trust" scales	Self report and questionnaires	In 119 measures, only 11 scales used in more than 1 study, and even then only for specific contexts

Behavior		
Facial features and expression	Rate trustworthiness of real or morphed faces	Rating of trustworthiness positively correlated with faces, but this was based on snapshots and no interactions – and no predictive power
Repeated "Trust" game w/ reputation	Earnings and response time	Trust led to increased earnings, decreased decision time
"Trust" game w/ prior information	Rate trustworthiness and earnings	Prior information led to increased monetary transfer

Neurophysiology		
fMRI	Wide variety of paradigms	Amygdala, Striatum, Precuneus, Anterior and Medial Cinculate Cortices, Insula, VTA, Septal Area, Paracingulate Cortex,
EEG	ERPs in evaluating faces	Distinct EEG waveforms in frontal electrodes when evaluating trustworthiness

Little understanding of contribution of different kinds of trust to behavioral outcomes

There has been a lack of cross-disciplinary collaboration, even in academia

Lack of repeated measures and definitions: Castaldo, 2004, found 72 definitions of trust, and McEvily & Tortoriello, 2007 found only 11 of 119 "trust" measures were actually repeated

Protocols are often low stress, anonymous, and even one-shot "games" with small consequences for betrayed trust

Protocols rely on "face" validity (har), but little construct validity

Still a long, long way to go

COGNITIVE NEUROSCIENCE
- Neuropeptides and hormones
- Neurophysiology involved in trust(s)
- New investigative tools

BEHAVIORAL ECONOMICS
- Increasing focus on operationalizing trust
- Death of "homo economicus"
- New intervention studies

SOCIAL SCIENCES and POLICY
- New appreciation for role of trust and culture
- Increased emphasis on emotionality
- Enhanced modeling techniques

SENSOR TECHNOLOGY
- Stand off capabilities
- Greater fidelity/resolution
- Multi-modal capabilities with better signal processing

Parallel breakthroughs in several key fields

Time is right for cross-disciplinary initiatives

People are not rational or irrational. They are *social* and need to be studied that way.
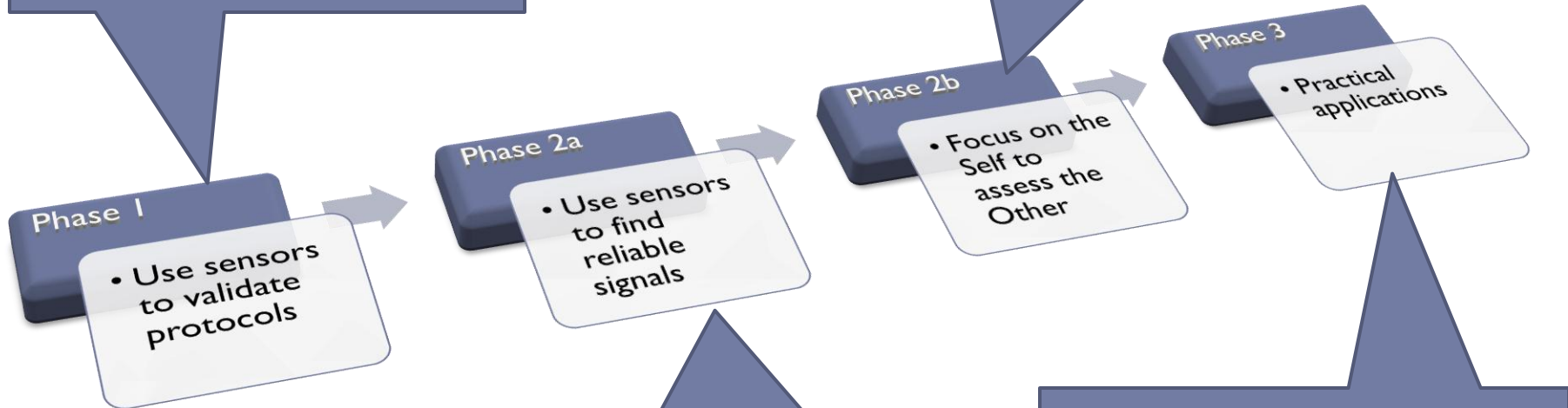
Program Overview

"**We need to stop** making what is measurable important, and find ways to make the important measurable."
- Robert McNamara

"**There are two possible outcomes**: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery."
- Enrico Fermi

Phase 1
- Use sensors to validate protocols

Phase 2a
- Use sensors to find reliable signals

Phase 2b
- Focus on the Self to assess the Other

Phase 3
- Practical applications

"**When you can measure** what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind."
- Lord Kelvin

"**Innovation is not** the product of logical thought, although the result is tied to logical structure."
- Albert Einstein

TRUST Fundamental Research Questions

Phase I
(This BAA)

▸ **Is there such a thing as "trust"? What are the different forms trust can take? What are the neurobiological processes by which trust emerges?**

- Can the signals indicative of these processes be quantified to characterize different types of trust for different people and contexts? The relative degree of trust and/or distrust?
- Can we experimentally disentangle trust from other social phenomena in a meaningfully, realistic, validated way?

Phase 2
(Potential future solicitations)

▸ **Can a Self's signals be a reliable, valid predictor of an Other's trustworthy behavior? Of the Other's decisions to trust?**

- Can supraconscious human assessment of trustworthiness be captured and processed in near real-time in order to assess who CAN be trusted?
- Rather than attempt to work around individual variability, can we leverage this variability to identify people who are highly predictive at detecting who will, and who will not, behave in a trustworthy manner?
- Can such assessments be reliable in specific individuals across critical human and contextual variables (language, culture, time, stress, etc.)?

Phase 3
(Potential future solicitations)

▸ **Can the self as a sensor be applied to real world dilemmas?**

Heisenberg Uncertainty Principle: Will the measurement of trust in another individual fundamentally alter natural interactions that foster trust within that interpersonal exchange?

By measuring it, have we changed it?

Will the prediction of a level of trust within an interaction enhance decision-making?

Notional Program Plan (Tentative)

- 5 years
- 3 Phases
- Multiple solicitations
- 1 Goal

- Envisioned to kick off in second quarter of FY 2010 and end in 2015

Phase 1 Overview

Phase 1 Goals

▸ Propose a model or models of trust based on literature review/pilot data you will test and (in)validate

▸ Propose specific hypotheses of what kinds of "signals" you expect to find across four domains if, indeed, your model(s) of trust have construct and ecological validity:

  ▸ Neural

  ▸ Physiological

  ▸ Psychological

  ▸ Behavioral

▸ Propose protocols **you will be able to run** (with sufficient power and incorporating key variables)

▸ Develop VALID protocols for assessing trust in realistic contexts (the TRUST Playbook)

  ▸ In Base Year (12 months) proposers must make sufficient and credible progress towards (in)validating their protocols to justify the award of the Option Year (12 months)

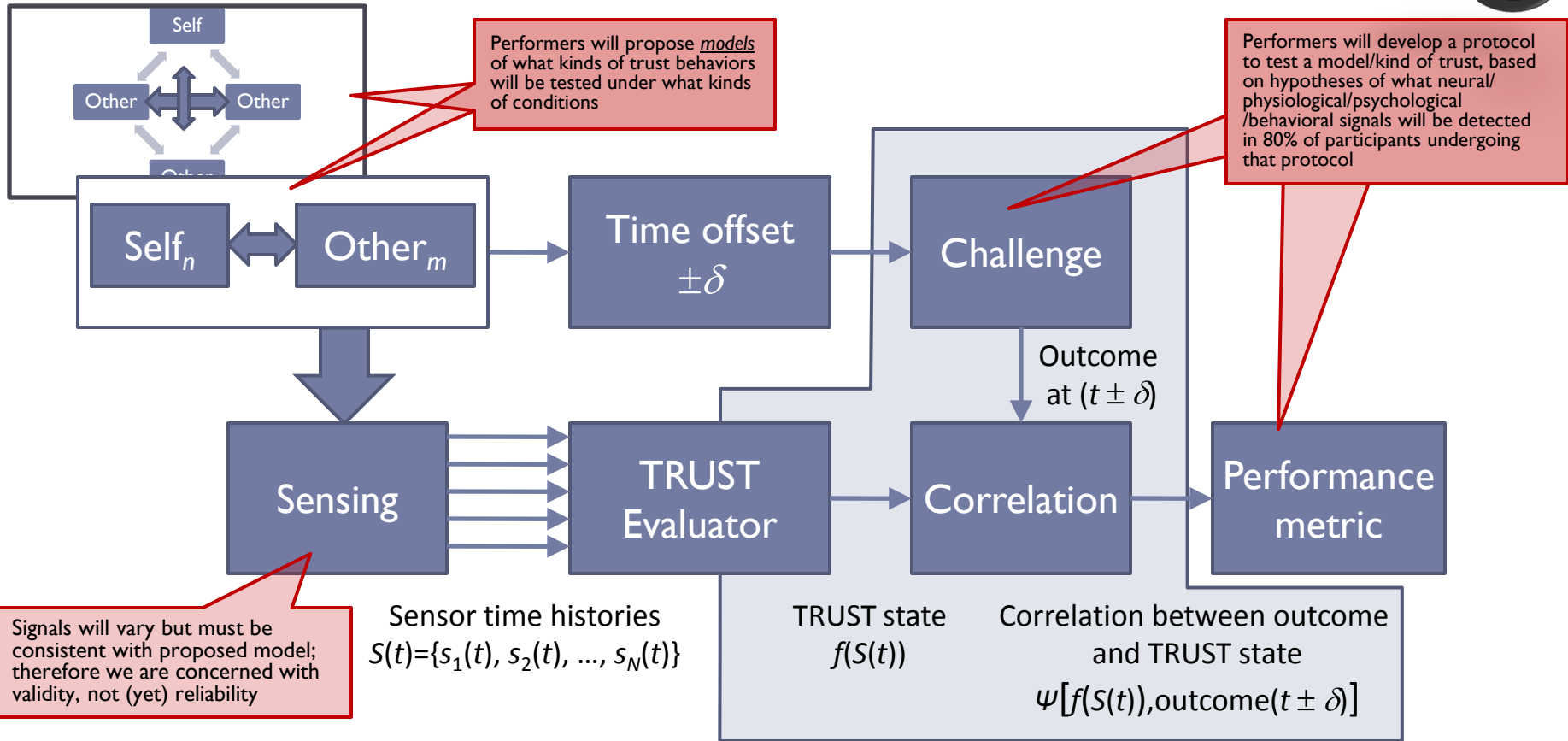 ▸ "Sufficient" and "credible" will be verified through the Program Manager and independent analysis/validation

Phase 1 Metrics

- Protocols MUST:
  - Be inter-cultural (proposers can define this but should justify it in proposals…and remember the PM is a social-cultural anthropologist)
  - Reflect "real world" conditions:
    - Time and stress
    - Dyads and small groups (proposers can define but should justify "small group")
    - Mixed genders/ages/ethnicities
    - Iterative
    - Non-anonymous
    - Credible motivation and high(er) risk/rewards
- Be VALID = face validity, construct validity, *and* ecological validity
  - Create protocols (what are you going to have people do to "test" trust?) based on your models of what kinds of trust are involved in different forms of interaction in dyads *and* small groups
  - Develop hypotheses to predict neural, physiological, behavioral, and psychological signals in subjects (0.80 validity across participants) that will either confirm or disprove the model and the protocol to a reasonably accurate degree
  - Collect data from all protocols ("discovery")
  - Collect validated protocols ("measurement")

Evaluation of Protocols
Phase 1

Performers will propose _models_ of what kinds of trust behaviors will be tested under what kinds of conditions

Performers will develop a protocol to test a model/kind of trust, based on hypotheses of what neural/ physiological/psychological /behavioral signals will be detected in 80% of participants undergoing that protocol

Self

Other ⟷ Other

Other

$Self_n$ ⟷ $Other_m$

Time offset $\pm\delta$

Challenge

Outcome at $(t \pm \delta)$

Sensing

TRUST Evaluator

Correlation

Performance metric

Signals will vary but must be consistent with proposed model; therefore we are concerned with validity, not (yet) reliability

Sensor time histories
$S(t)=\{s_1(t), s_2(t), ..., s_N(t)\}$

TRUST state
$f(S(t))$

Correlation between outcome and TRUST state
$\Psi[f(S(t)),\text{outcome}(t \pm \delta)]$

Phase 1: Prove validity of TRUST protocol/model by developing and testing hypotheses to predict signals across 0.8 of subjects in each of four domains: Neural, psychological, physiological, behavioral

Examples

▸ IF this protocol actually tests trust as we model it…

▸ THEN we hypothesize that we shall see these neural/psychological/physiological/behavioral signals that are consistent with this model

  ▸ Signals will obviously differ among participants as some will or will not trust (may not be reliable)

  ▸ Regardless of who does or does not prove to be trustworthy, signals across > 0.80 of participants will be consistent with the model and thus validate the protocol as testing trust under those conditions (as opposed to testing risk taking, or boredom, or spite for the researchers, or having nothing else to do, or antipathy, etc.)

Phase 1 Metrics

METRICS	Phase I
Models for Dyad and Small Group	Pass
Protocols address key variables for Dyad and Small Group	Pass
Predicted Neural Signal(s)	0.80 across all subjects in a specific protocol
Predicted Behavioral Signal(s)	0.80 across all subjects in a specific protocol
Predicted Physiological Signal(s)	0.80 across all subjects in a specific protocol
Predicted Psychological/Psychometric Signal(s)	0.80 across all subjects in a specific protocol
Signals are consistent with models and validate protocols	Pass

Research Method Considerations

▸ Must provide own laboratory facilities

▸ Specify (and justify):

   ▸ Key variables

   ▸ Number of human subjects in small groups

   ▸ Planned length of the data gathering period

   ▸ Model(s)

   ▸ Hypotheses

▸ Qualitative research methods

   ▸ Qualitative methods may be used as an adjunct, to inform the direction of the quantitative research, or to help interpret the results of the quantitative research.

▸ Sample size

▸ Control to avoid biasing own results

UNCLASSIFIED

More Research Method Considerations

- Technique(s) for safeguarding privacy
  - Safeguard data sets from accidental release or malicious intrusions
  - IARPA reserves the right to reject proposals that do not adequately address the safeguarding of privacy

- Must include appropriate techniques for safeguarding Personally Identifiable Information (PII)

- IARPA will <u>not</u> receive any raw data

- The TRUST Program will have independent validation teams from NIST and other USG organizations
  - Data set access may be provided to independent validators as needed

- Independent validators will retain data access no longer than 12 months after the conclusion of Period of Performance

See the section on Civil Liberties and Privacy Protection Measures for more details

Other Research Proposal Considerations

▶ Whitepaper period is expected (you may like to submit Organizational Conflict of Interest paperwork with any whitepaper, although this is not mandatory)

▶ You can find OCI policy at
http://www.iarpa.gov/IARPA_OCI_081809.pdf

See section on addressing Organizational Conflict of Interest – and read carefully

Out of Scope

▸ Detecting deception

▸ Buying your very own fMRI and MEG

▸ Putting lipstick on the "trust game"

▸ Fishing expeditions

▸ Meta-analyses

▸ Social network analysis

▸ Disease or pathology

▸ Tangentially related concepts like "empathy"

▸ Development of individual novel sensor technology

UNCLASSIFIED

Questions?

UNCLASSIFIED

Phase 2 & 3 Overview

Phase 2

- Use validated protocols and sensing to assess reliability of "trust" signals in the Self and the Other

  - In dyad and small group, focus is on Self's signals in first part of the protocol to assess (forecast) the Other's/Others' trustworthiness in second part of the protocol

    - Intercultural (does not have to mean international)
    - Training and testing samples to allow individual calibration
    - Combination of signals from Self *and* Other may be collected
    - Appropriate "controls" will be important

  - Performers will select as many sensors as they like, but must be at or close to COTS (although they can be used in novel ways)

  - Algorithms and combinations of sensors cannot be proprietary as they will be used in Phase 3

  - Two collection Phases:

    - **Phase 2a: ("Sensorama")** Any combination of sensors that can be used without deviating from protocols – find the signals and accuracy may be determined after the fact ("retro-dictive")

    - **Phase 2b: ("Hands Off")** Only sensors that can be used in near real-time, and only on the Self to find signals that assess who will be trustworthy ("pre-dictive")

Phase 2a

▸ **Measure Self and Other during validated protocol using combinations of sensors**

is prediction accurate?

measured behavior   *predicted* behavior

Other

Validated Protocol
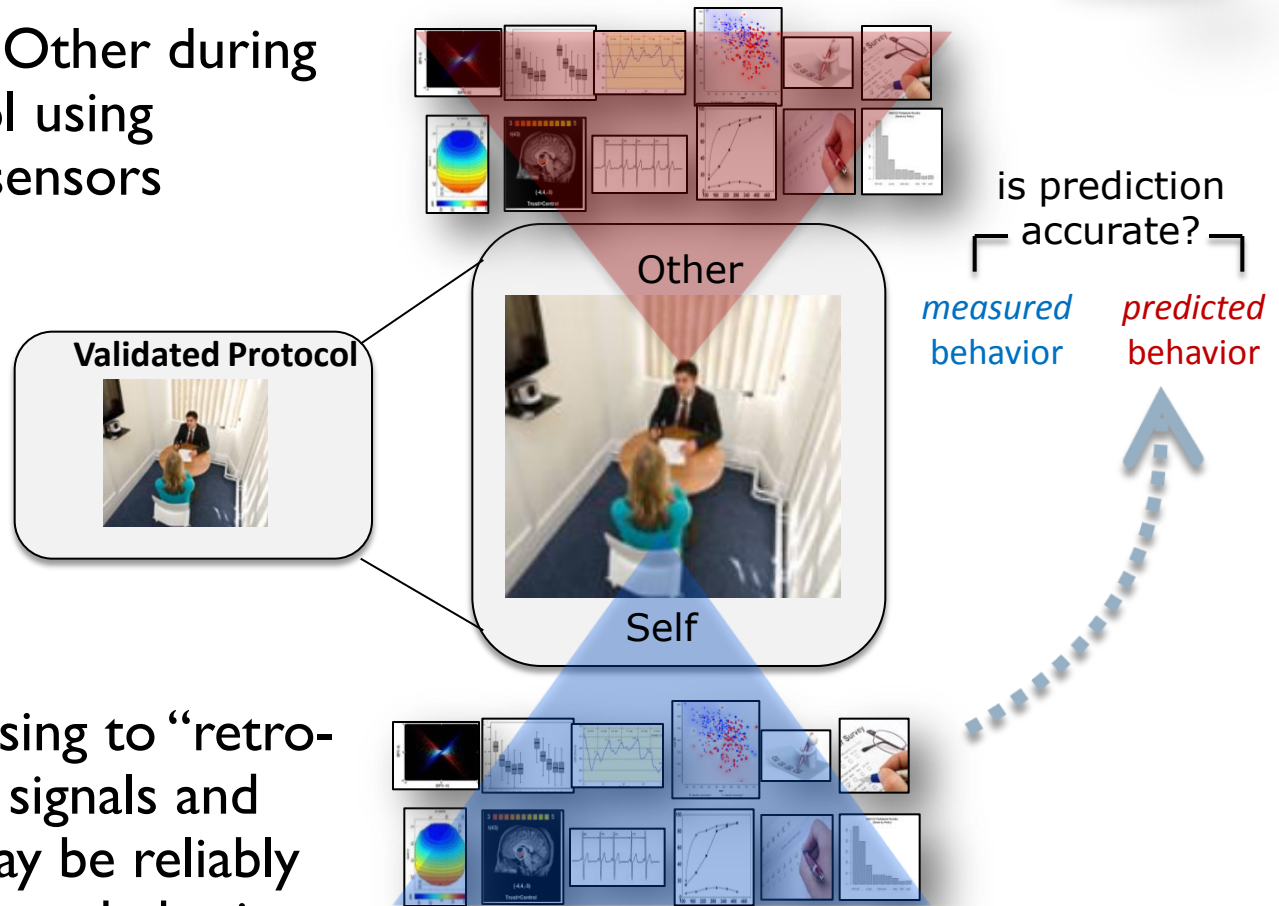
Self

▸ **Use signal processing to "retro-dictively" identify signals and signatures that may be reliably correlated with trust behaviors during protocol**

Phase 2b

▸ Use sensors only on Self to assess individual signatures during validated trust protocols for "*p*redictive" reliability
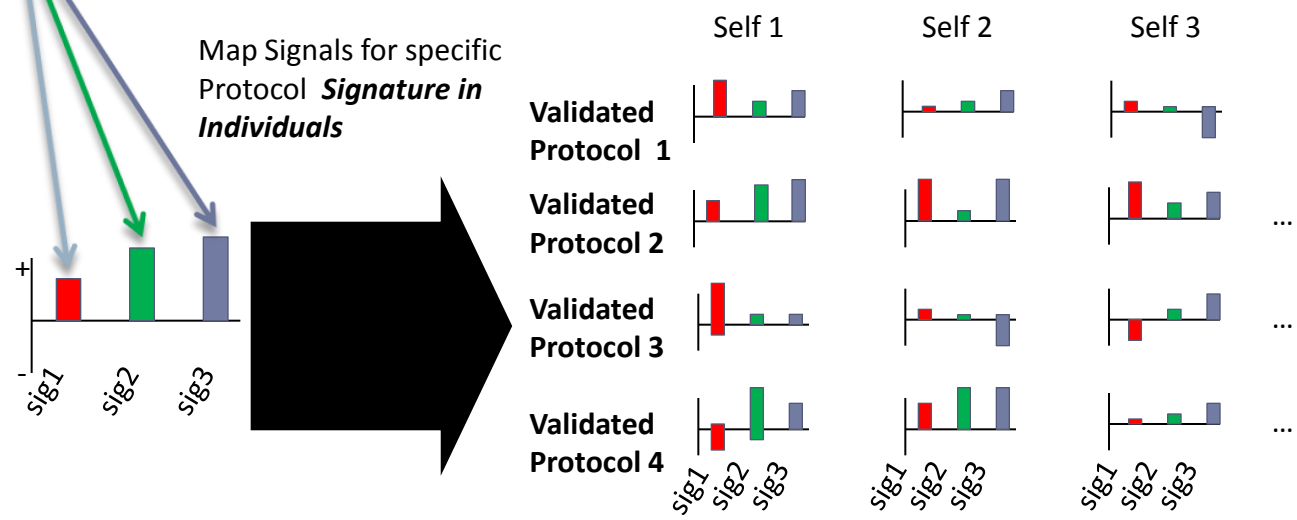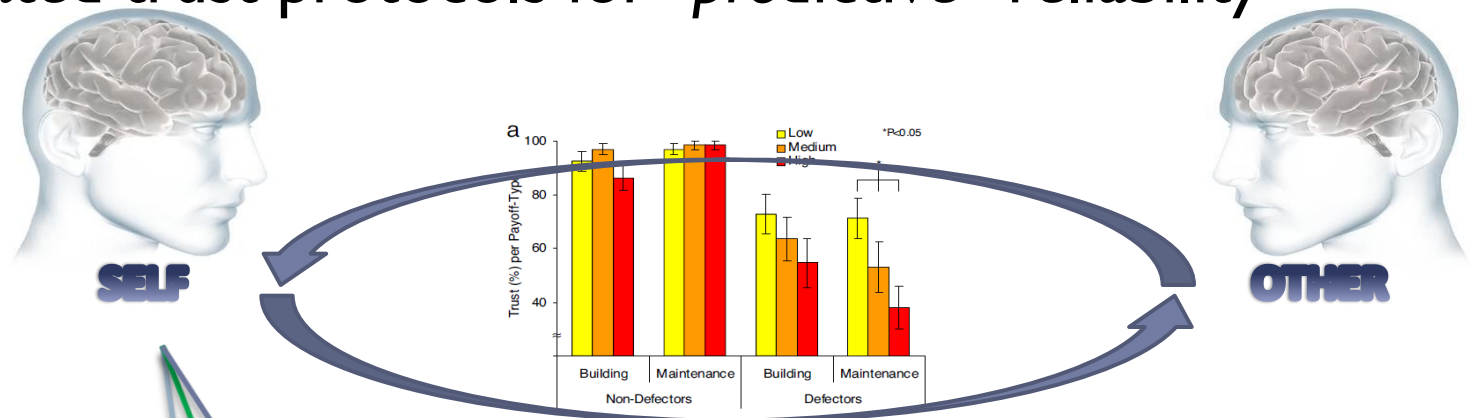
Phase 3

- Bring most promising sensors and algorithms into practical environments for independent validation with USG team

- Sensors/algorithms calibrated to individuals

- Sensors must:
  - Provide near-real time assessment
  - User-friendly interface ("TRUST dashboard")
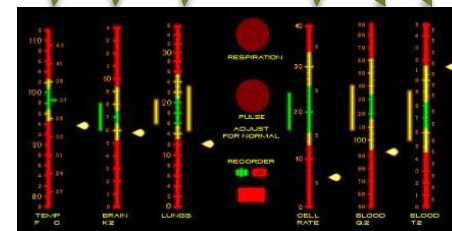
Questions?

Programmatics

UNCLASSIFIED

Civil Liberties and Privacy Protection Measures

▸ Must address the following (as applicable):

  ▸ Comply with federal policy for protection of human subjects in research

  ▸ Receive approval from their Institutional Review Board

  ▸ Obtain informed consent when using intervention and/or interaction

  ▸ Employ techniques to protect privacy and confidentiality

There will be an annual review by TRUST PM and ODNI Civil Liberties and Privacy Office

Human Subjects

- Must provide evidence of, or a plan for, review by an IRB upon final proposal submission to IARPA

- Allot ample time to complete the approval process

- IARPA will not possess or have access to ANY individually identifiable information

- No IARPA funding can be used towards human subjects research until ALL approvals are granted

IARPA Proposal Submissions

1. Download ALL materials posted to the web site (BAA, instructions, templates, etc.)

2. Periodically check for amendments and other information that may be posted prior to the proposal due date

3. Read FAQs posted to the web site

4. Ensure submission requirements are followed:

   - Deadlines

   - Do not exceed page limits

   - Use all provided templates (see Appendix)

   - Include all required responses (OCI paperwork, Academic Acknowledgement letters, etc.) – might want to include OCI paperwork with whitepaper period

   - No unnecessarily elaborate brochures or marketing material

   - Failure to follow the submission procedures may result in the submission not being evaluated.

Evaluation Criteria

Evaluation criteria are:

▸ Overall Scientific and Technical Merit

▸ Effectiveness of Proposed Work Plan

▸ Relevance to IARPA Mission and TRUST Program Goals

▸ Relevant Experience and Expertise

▸ Cost Realism

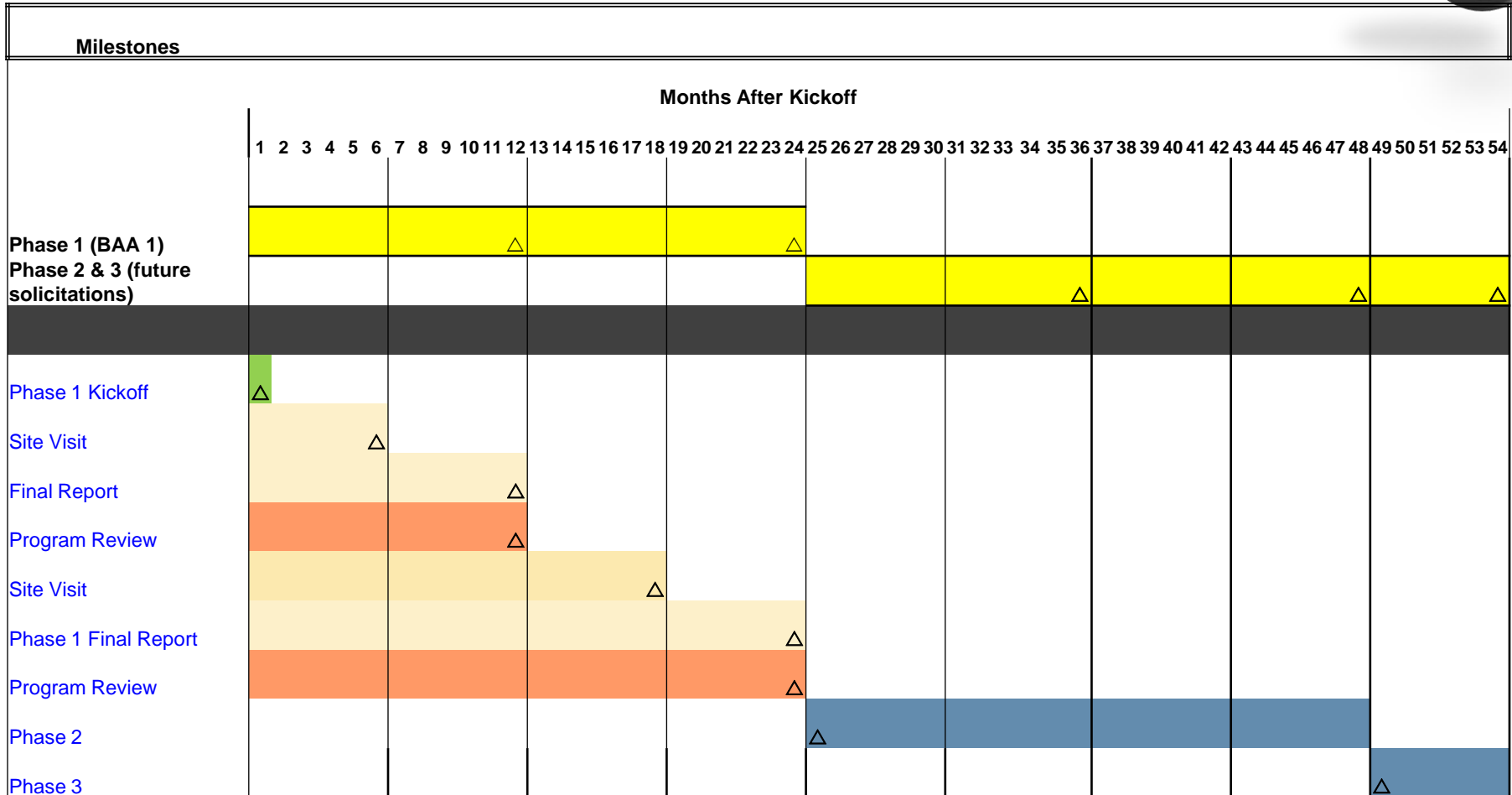# Award Plan

- BAA 1
  - Phase 1 (24 months)
    - Base Period – 12 months
    - Option Period – 12 months

- Phase 2 & 3 expected to follow as:
  - 24 months (Phase 2)
  - 6-12 months (Phase 3)

- Multiple awards anticipated, depending upon:
  - Quality of the proposals received
  - Availability of funds

Milestones

Milestones																																																					

Months After Kickoff

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Phase 1 (BAA 1)											△												△																															
Phase 2 & 3 (future solicitations)																																			△									△				△						△
Phase 1 Kickoff	△																																																					
Site Visit						△																																																
Final Report												△																																										
Program Review												△																																										
Site Visit																		△																																				
Phase 1 Final Report																							△																															
Program Review																							△																															
Phase 2																									△																													
Phase 3																																																		△					

Offerors should suggest additional 6 month metrics/milestones that will enable the PM to assess their progress

Reporting

- ▸ Minimum requirement: monthly technical and financial status reports

- ▸ Final report at conclusion of Base and Option years

- ▸ Demonstrations at Program Reviews as appropriate

BAA 1 / Phase 1

Requirements:
- Validated protocols- 0.80 minimum probability of detecting specific type of trust in protocols
- Report performance metrics (behavior, psychometrics, neuro, physiology) and detailed methodologies
- File formats Matlab compatible, MAT-file (EDF, CDF, XML, etc., etc.)
- Other groups should be able to read your data
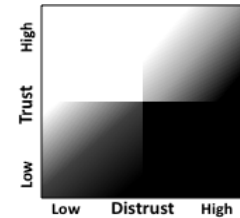
Meeting and Travel Requirements

- In each Phase, plan for Program Kick-off meetings and Program Reviews

  - in Metropolitan DC area

  - to facilitate an open exchange among all Program participants

- Each meeting 2-3 days

- Periodic site visits from PM and USG support as appropriate

Eligible Applicants

- Did we mention that collaborative efforts/teaming strongly encouraged???
    - Content, communications, networking, and team formation - responsibility of proposers
- Foreign organizations and/or individuals may participate
    - Must comply with Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc, as appropriate

Ineligible Organizations

▸ The following are **NOT** eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities:

  ▸ Other Government Agencies

  ▸ Federally Funded Research and Development Centers (FFRDCs)

  ▸ University Affiliated Research Centers (UARCs)

  ▸ Any organizations that have a special relationship with the Government, including access to privileged and/or proprietary information, or access to Government equipment or real property.

UNCLASSIFIED

Organizational Conflict of Interest

▸ If a prospective offeror, or any of its proposed subcontractor teammates, believes that a potential conflict of interest exists or may exist (whether organizational or otherwise), the offeror should promptly raise the issue with IARPA and submit a waiver request by e-mail to the mailbox address for this BAA at dni-iarpa-baa-10-03@ugov.gov. A potential conflict of interest includes - ***but is not limited to*** - any instance where an offeror, or any of its proposed subcontractor teammates, is providing either scientific, engineering and technical assistance (SETA) or technical consultation to IARPA. In all cases, the offeror shall identify the contract under which the SETA or consultant support is being provided. Without a waiver from the IARPA Director, neither an offeror, nor its proposed subcontractor teammates, can simultaneously provide SETA support or technical consultation to IARPA and compete or perform as a Performer under this solicitation.

▸ Please read the ***entire*** OCI policy at:
http://www.iarpa.gov/IARPA_OCI_081809.pdf

Publication Approval

‣ You are encouraged to publish (in fact, if you're not publishing, we will want to know why)

‣ BUT…

‣ Pre-publication approval may be required for sensitive information

‣ Any award may include a requirement for pre-pub review

‣ It is also considered appropriate for performers to provide a soft copy to:

   ‣ the IARPA TRUST Program Manager and

   ‣ the Contracting Officer Representative (COR)

Point of Contact

Dr. Adam H. Russell

Program Manager

IARPA, Smart Collection Office

Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity
Washington, DC 20511

Phone: 301-226-9100
Fax: 301-226-9137

Electronic mail: dni-iarpa-baa-10-03@ugov.gov
(include IARPA-BAA-10-03 in the Subject Line)

Website: www.iarpa.gov

UNCLASSIFIED

Thanks!

UNCLASSIFIED

Backups

UNCLASSIFIED

Funding Opportunity Description

▸ Starting from the premise that people have neurophysiological, behavioral, and psychological signals when in the presence of others who are trustworthy, the IARPA TRUST program seeks to develop tools that can validate, detect, and assess such signals of trustworthiness.

▸ The TRUST BAA will appear first on the FedBizOpps website (http://www.fedbizopps.gov/)

▸ Q&As will appear after the BAA

Overview Information

- The TRUST Broad Agency Announcement (BAA) sets forth research areas of interest in the area of sensing to detect and assess interpersonal trust and trustworthiness

- Awards based on responses to this BAA are considered to be the result of full and open competition.

- Federal Agency Name – Intelligence Advanced Research Projects Activity (IARPA), Incisive Analysis Office

- Funding Opportunity Title – TRUST

- Announcement Type – Initial

- Funding Opportunity Number – IARPA-BAA-10-03

- Catalog of Federal Domestic Assistance Numbers (CFDA) –Not applicable.

- Dates: Proposal Due Date: (Insert proposal due date)

- Concise description of funding opportunity: Starting from the premise that people have neurophysiological, behavioral, and psychological signals when in the presence of others who are trustworthy, the IARPA TRUST program seeks to develop tools that can validate, detect, and assess such signals of trustworthiness.

- Anticipated individual awards – Multiple awards are anticipated.

- Type of instruments that may be awarded – Procurement contract.

Export Control

▶ Research teams shall comply with all U.S. export control laws and regulations, including the International Traffic in Arms Regulations (ITAR), 22 CFR Parts 120 through 130, and the Export Administration Regulations (EAR), 15 CFR Parts 730 through 799, in the performance of this contract

▶ Research teams shall be responsible for obtaining required export licenses <u>before</u> using foreign persons in the performance of this contract,

    ▶ Including instances where the work is to be performed on-site at any Government installation (whether in or outside the United States) or,

    ▶ Where the foreign person will have access to export-controlled technologies, including technical data or software

▶ Each research team shall be responsible for:

    ▶ All regulatory record keeping requirements associated with the use of licenses and license exemptions/exceptions

    ▶ Ensuring that the provisions of this clause apply to its subcontractors

▶ Research teams will certify knowledge of and intended adherence to these requirements in the representations and certifications of the contract

UNCLASSIFIED

Human Subjects

❑   IARPA considers research interpersonal trust and trustworthiness to involve human subjects

▸ Research involving human subjects that is conducted or supported by the DoD must comply with:

  ▸ 32 CFR 219, *Protection of Human Subjects,* http://www.dtic.mil/biosys/downloads/32cfr219.pdf

  ▸ DOD Directive 3216.02 *Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research,* http://www.dtic.mil/whs/directives/corres/pdf/321602p.pdf.

▸ Institutions awarded funding for research involving human subjects must provide documentation of a current Assurance of Compliance with Federal regulations for human subject protection,

  ▸ For example see Department of Health and Human Services, Office of Human Research Protection Federal Wide Assurance (http://www.hhs.gov/ohrp)

  ▸ All institutions engaged in human subject research, to include subcontractors, must also have a valid certification of compliance