

INfluence Campaign Awareness & Sensemaking (INCAS)

Brian Kettler, PhD
I2O

Proposers Day

30 October 2020





Agenda

Start	End	Duration	Item
12:00 Noon	1:00 PM	1:00	Online Registration
1:00 PM	1:05 PM	0:05	Security Briefing Mr. Christian Seth
1:05 PM	1:20 PM	0:15	Human Use Briefing Ms. Lisa Mattocks, I2O ADPM/HSR Action Officer
1:20 PM	1:45 PM	0:25	Contracts Management Office Briefing Ms. Jennifer Mack, Contracting Officer
1:45 PM	2:30 PM	0:45	Influence Campaign Awareness and Sensemaking (INCAS) Presentation Dr. Brian Kettler, Program Manager, DARPA I2O
2:30 PM	2:35 PM	0:05	Submit Questions
2:35 PM	4:00 PM	1:25	PM Question Review and Informal Team Discussions
4:00 PM	4:45PM	0:45	Q&A Session (Answer attendee questions)



Key Information

- BAA Locations and Dates
 - Posted on FedBizOpps website (<https://fbohome.sam.gov/>)
 - Posting Date: October 26, 2020
 - Abstract Due Date: November 17, 12:00 noon (ET)
 - BAA Closing (Proposal Due Date): January 8, 2020, 12:00 noon (ET)
- Procedure for Questions/Answers Today
 - Questions can be submitted until 2:35PM (ET) to INCAS@darpa.mil
 - Questions will be answered during Q&A session in the afternoon
 - Questions that are not addressed during the Q&A session will be posted on FAQ site
- Websites
 - Proposers' Day website: <https://www.schafertmd.com/DARPA/I2O/INCAS/PD/2020/Oct/>
 - INCAS program website
 - Proposers Day Slides
 - Frequently Asked Questions (FAQ) will be updated with Q/A from INCAS@darpa.mil



Objective

Develop analyst-guided techniques and tools to detect and track geopolitical influence campaigns with quantified confidence.



Challenges

- The US is engaged with its adversaries in an asymmetric, continual, war of weaponized influence narratives.
 - Adversaries exploit misinformation and true information delivered via influence messaging: blogs, tweets, and other online multimedia content.
 - Princeton's Empirical Studies of Conflict Project document lists nearly 100 foreign and domestic influence campaigns in the past 9 years.
 - Online influence campaigns can have real-world (offline) outcomes.
- Analysts require effective tools for continual sensemaking of the vast, noisy, adaptive information environment to identify adversary influence campaigns. Focus is on sensemaking, not countering influence operations.
- INCAS is an applied research and development effort and is thus expected to result in portable, modular tools and technologies that operational users can assess.
- INCAS tools will be demonstrated and evaluated on publicly available data for several historical and current scenarios of potential adversary campaigns in non-U.S. populations.
- INCAS will engage with operational stakeholders over the course of the program for continual feedback on tools and to help position tools for transition.
- INCAS research and technology development will be unclassified.

*This report uses only *open source* media reports and research articles.

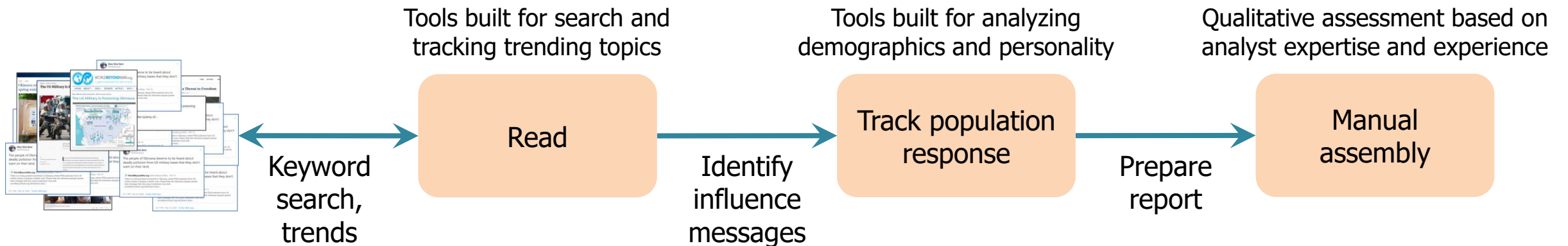
<https://esoc.princeton.edu/publications/trends-online-influence-efforts>



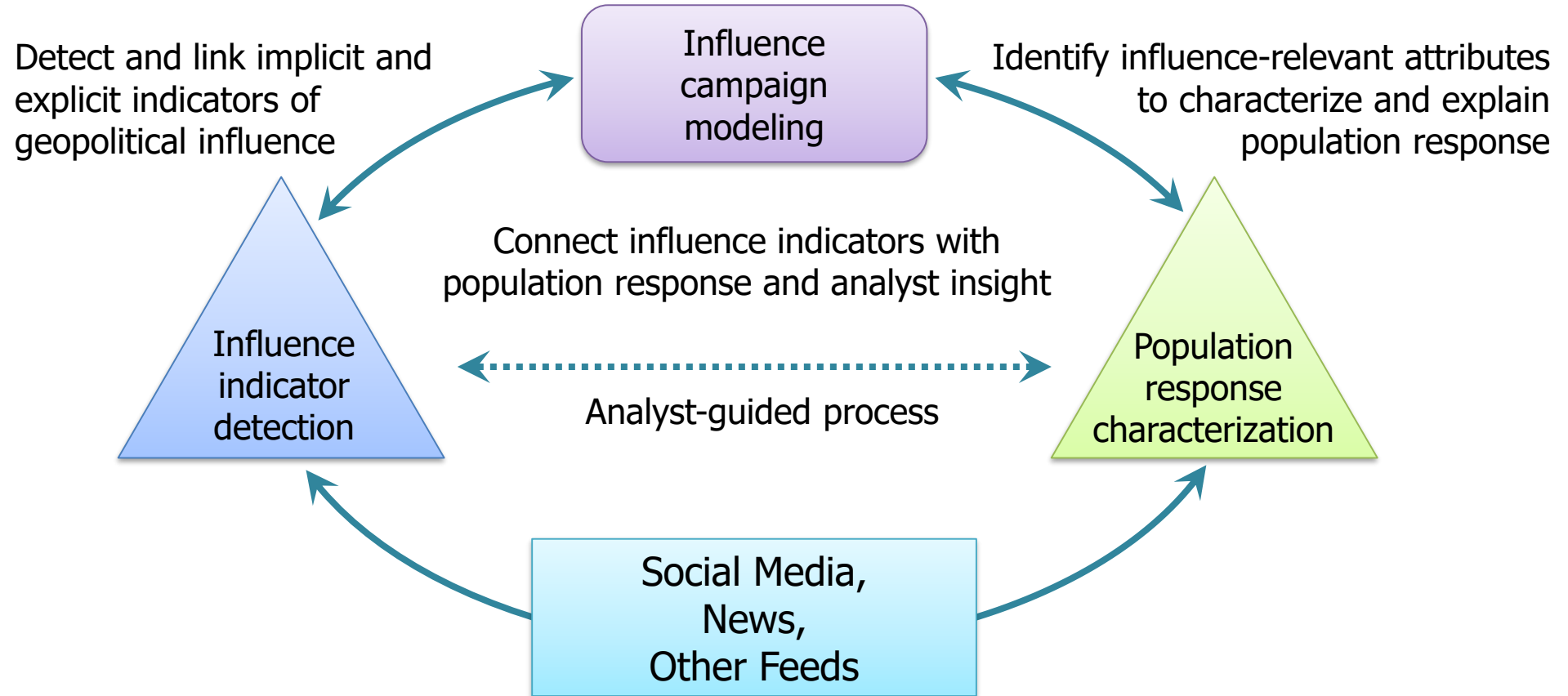
How influence campaign detection and sensemaking are done today

Current techniques are ad hoc, manual, slow, and lacking principled confidence assessment

- Analysts must formulate complex keyword queries; track trending keywords; and read hundreds or thousands of documents to identify influence themes
- Analysts track population response using tools limited to analyzing demographics and personality
- Analysts have difficulty connecting messaging over time and across multiple platforms to see evolving campaigns
- Analysts have difficulty assessing confidence in analytic conclusions due to varying expertise, experience, and biases



Analyst-guided campaign analysis using automated and measured influence detection

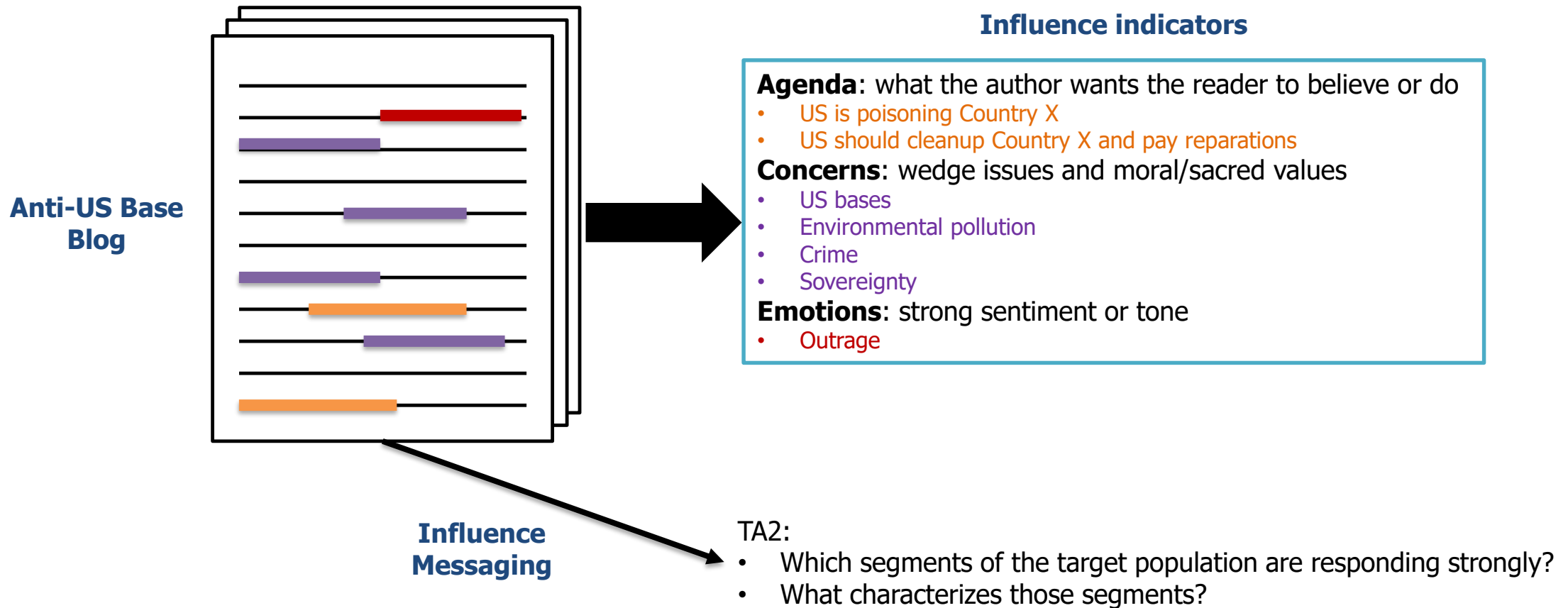




Indicators of geopolitical influence

Today: Current social media tools fail to detect emerging and implicit geopolitical influence indicators in messaging

New Approach: Automatically detect implicit and explicit indicators of geopolitical influence in multilingual online messaging

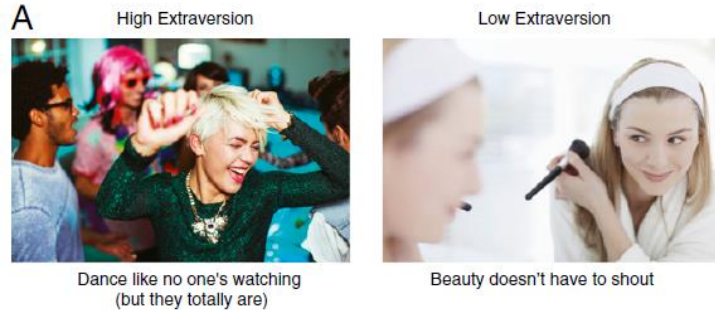




Population response characterization

Today: Populations are segmented and characterized based on pre-defined categories (e.g., demographics) & personality traits using techniques and tools developed for marketing

New Approach: Dynamically segment and characterize populations based on their response to influence messaging using novel psychographic attributes (e.g., worldviews, values)



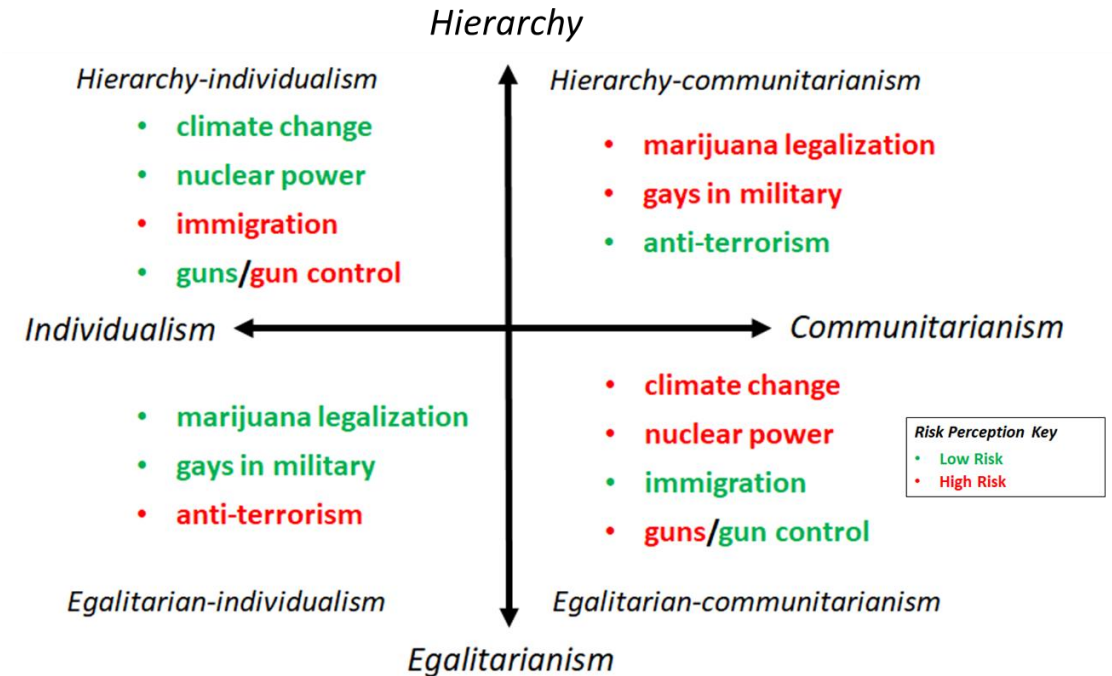
Beauty ads designed for extroverted women (left) and introverted women (right)



Game app ads designed for people with high openness to new experience (highlight novelty) and low openness (highlight familiarity)

Marketing: Ad targeting using demographic and personality attributes increased click-thru by 40% and purchases by 50%.

S. C. Matz, M. Kosinski, G. Nave and D. J. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion," *Proc. of the Natl Academy of Sciences*, vol. 114, no. 48, 2017.



Psychographic attributes such as worldviews (e.g., attitudes towards hierarchy and individualism) impact perception of social concerns.

Yale Cultural Cognition project: <http://www.culturalcognition.net/blog/2018/1/6/culture-worldviews-risk-perception-glossary-entries.html>, originally from: D. M. Kahan, "Cultural cognition as a conception of the cultural theory of risk," in *Handbook of Risk Theory*, 2008.



Influence campaign modeling

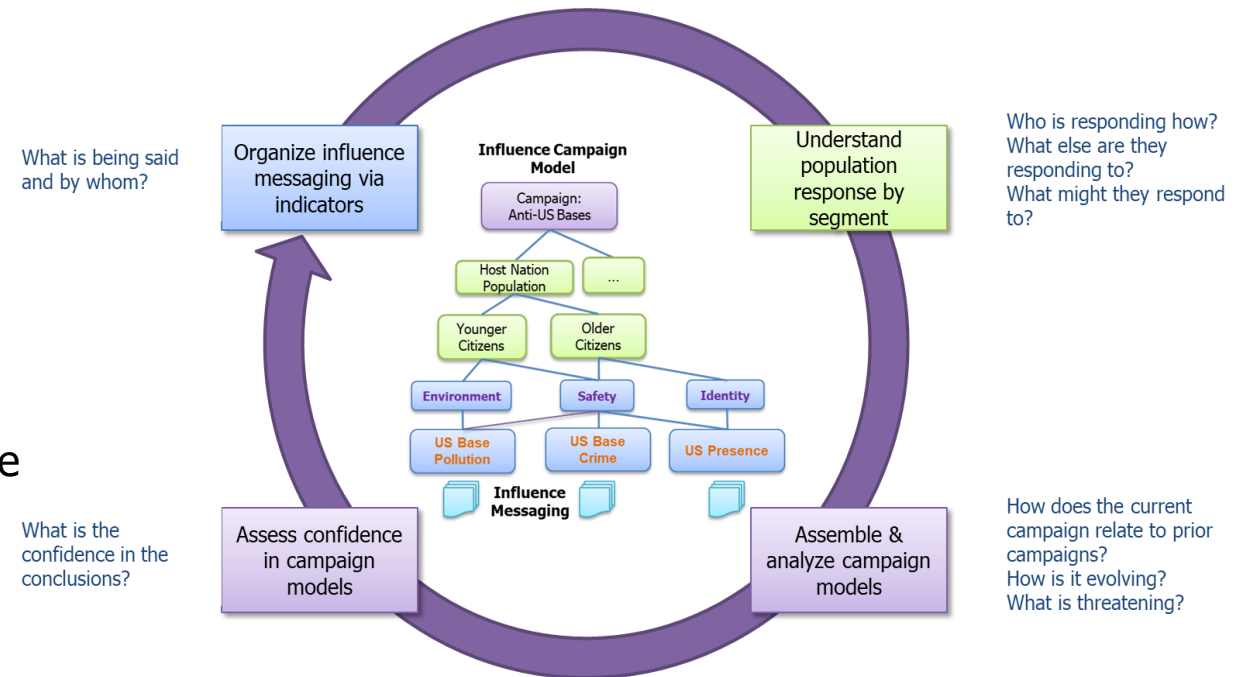
Today:

- Analysts organize messaging via link analysis tools and integrate related intelligence to produce a static report
- Confidence assessment by analysts is ad hoc, manual, subjective, qualitative and susceptible to analyst biases

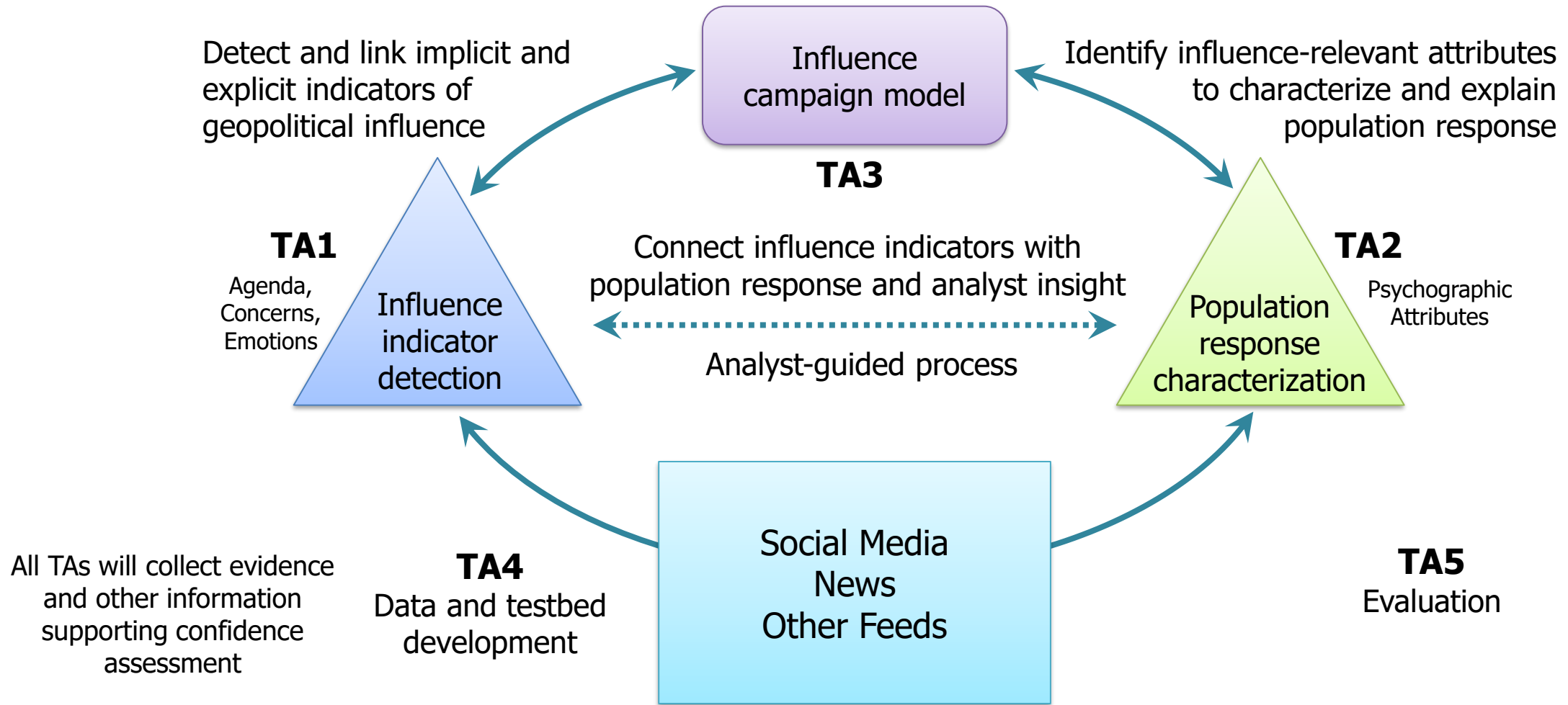
New Approach:

- Analyst-guided campaign modeling tools accelerate analysts' ability to detect/link influence and response over time
- Machine curates, elicits, combines, and organizes information for analysts to quantitatively assess confidence in campaign models
- **Campaign models** link machine-surfaced and analyst-provided elements over time:
 - influence indicators and messaging within and across platforms
 - population response and psychographic attributes
 - campaign targets, tactics, objectives, actors, events, etc.

Analyst-guided campaign modeling (sample workflow)

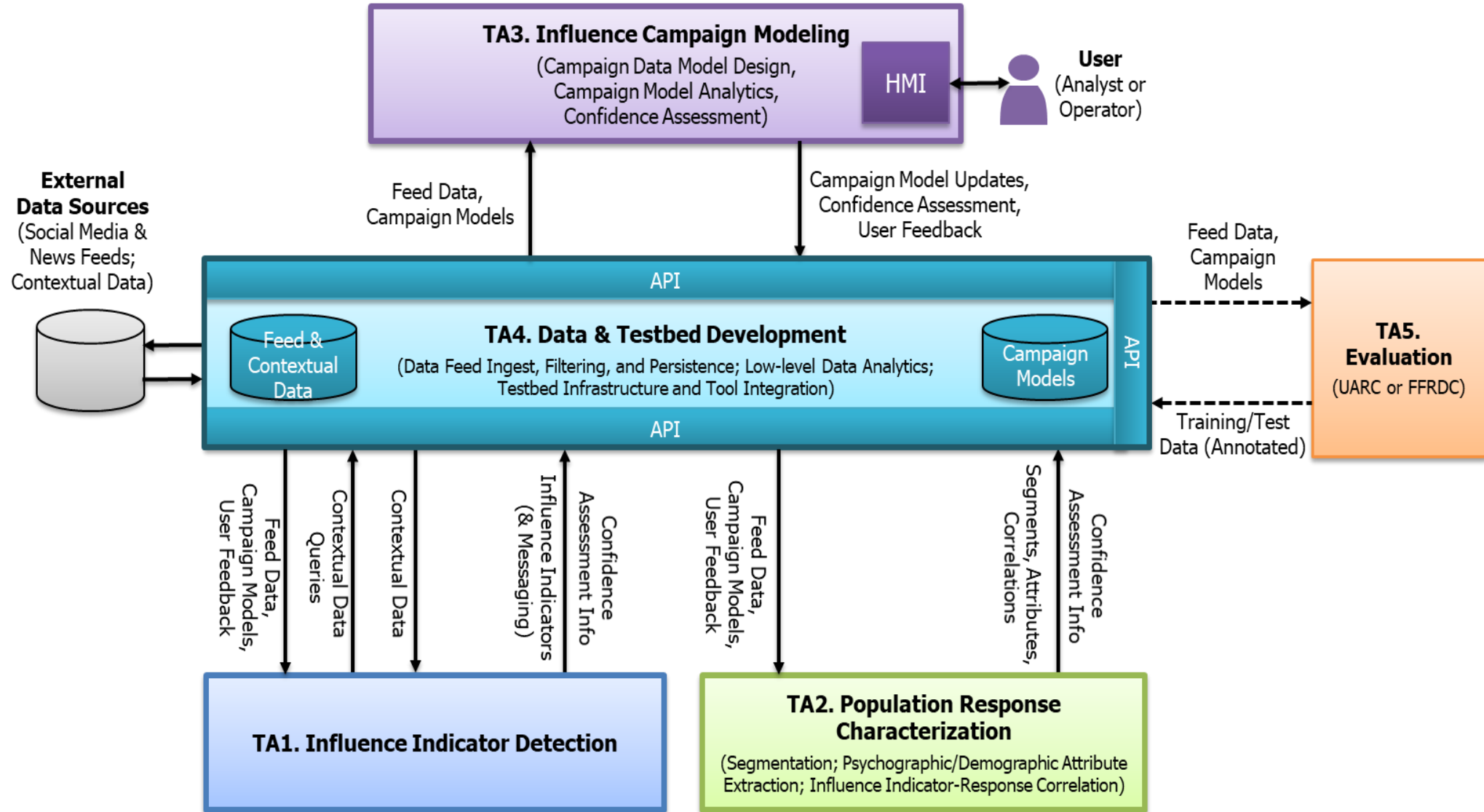


Analyst-guided campaign analysis using automated and measured influence detection





Technical areas: data flows





Influence indicator detection (TA1)

Goal

- Automatically detect implicit and explicit indicators of geopolitical influence in multilingual online media to include, but not limited to:
 - Strong emotion/sentiment (e.g., outrage, despair)
 - Deeper concerns: wedge issues and moral/sacred values
 - Agenda: what the author wants the reader to believe or do
 - Other indicator types (2 in Phase 1, 2 in Phase 2) exploiting content, metadata, or other (structural, temporal, etc.)

Technical Challenges

- Extract explicit and implicit influence indicators at speed/scale across a broad range of geopolitically-relevant domains of discourse
- Handle fragmentary and implicit text
- Handle multilingual and multicultural text (English + 2 languages)
- Exploit campaign model context and user feedback (via TA3)
- Provide confidence assessment inputs (to TA3)

Metrics

- Influence messaging detection accuracy (P_D and P_{FA})
- Indicator extraction accuracy

Describe

- For *each* indicator type: representation, extraction approach, examples, theoretical/empirical basis, utility/generalizability
- Use/source of contextual data
- Use/source of training data



Population response characterization (TA2)

Goal

- Dynamically segment populations based on their response to influence and characterize the response using demographic and psychographic attributes (e.g., worldviews)

Technical Challenges

- Segment responding population at sufficient granularity: e.g., based on online actions, emotional response, etc. (data-driven)
- Extract psychographic attributes / behavioral patterns from online data that have explanatory/predictive power, analyst interpretability, cultural generality, and longevity
- Correlate with influence indicators to explain/predict response
- Exploit campaign model context and user feedback (via TA3)
- Provide confidence assessment inputs (to TA3)

Metrics

- Psychographic/demographic attribute identification accuracy
- Correlation of population attributes with influence indicator accuracy

Describe

- 2 demographic and 2 psychographic attributes in each of Phase 1 and Phase 2
 - For *each* attribute: relevance to geopolitical influence campaigns; extraction approach; theoretical support from scientific/marketing literature; expected utility; generality across different populations, cultures, languages and over time; and feasibility of extraction from publicly available online data.
 - Analyst interpretability (for patterns)
- Use/source of training data (and mitigation of cognitive biases – e.g., survey data)
- Segmentation and correlation approaches



Influence campaign modeling (TA3)

Goals

- Analyst-guided campaign modeling tools accelerate analysts' ability to detect and link influence and response over time
- Machine curates, elicits, combines, and surfaces confidence assessment information & evidence to mitigate biases

Technical Challenges

- Enable humans and machines to jointly and iteratively model large, evolving campaigns
 - Link influence indicators, population response, and other campaign elements (e.g., actors, objectives, tactics, events) over time, across platforms
 - Scale to 100s-1000s of concurrent models, 0.1K-10K elements each
 - Assist analysts in assessing campaign origin, threat
- Aggregate and present information to enable analysts to continually make quantitative confidence assessments of campaign models
 - Handle multiple sources of potential bias/error (e.g.: data, algorithms, analyst)
 - Combine heterogeneous confidence assessment information from TA1-TA4, including information elicited from analysts

Metrics

- Sensemaking scores (analyst questionnaires) for analysts with and without INCAS tools
- Utility & usability of confidence assessment information (SME judged)

Describe

- Human-machine interaction / interface (usability, flexibility, scalability, etc.)
- Campaign data model design approach (TA4 to implement supporting services)
- Analytics (triage, filter, cluster, etc.): algorithms, data, validation
- Campaign confidence assessment approach
- Plan to work with users (requires some personnel cleared at least to SECRET level)



Data and testbed development (TA4)

Goals

- Provision multimedia data for tool development and evaluation
- Enable tools to be easily assimilated by operational users

Technical Challenges

- Ingest multilingual, multicultural, and multimedia data
 - text focus but investigate images (ads, memes)
- Process images (ads, memes) to obtain descriptive text
- Filter volume to most relevant messaging based on target population, evolving campaign model, and analyst feedback
- Collect and exploit relevant metadata (e.g., cyberforensic indicators (bots, falsified media), hashtags, emoticons/emoji)
- Develop extensible infrastructure for tool integration, test & evaluation, and deployment (individually or in combination)

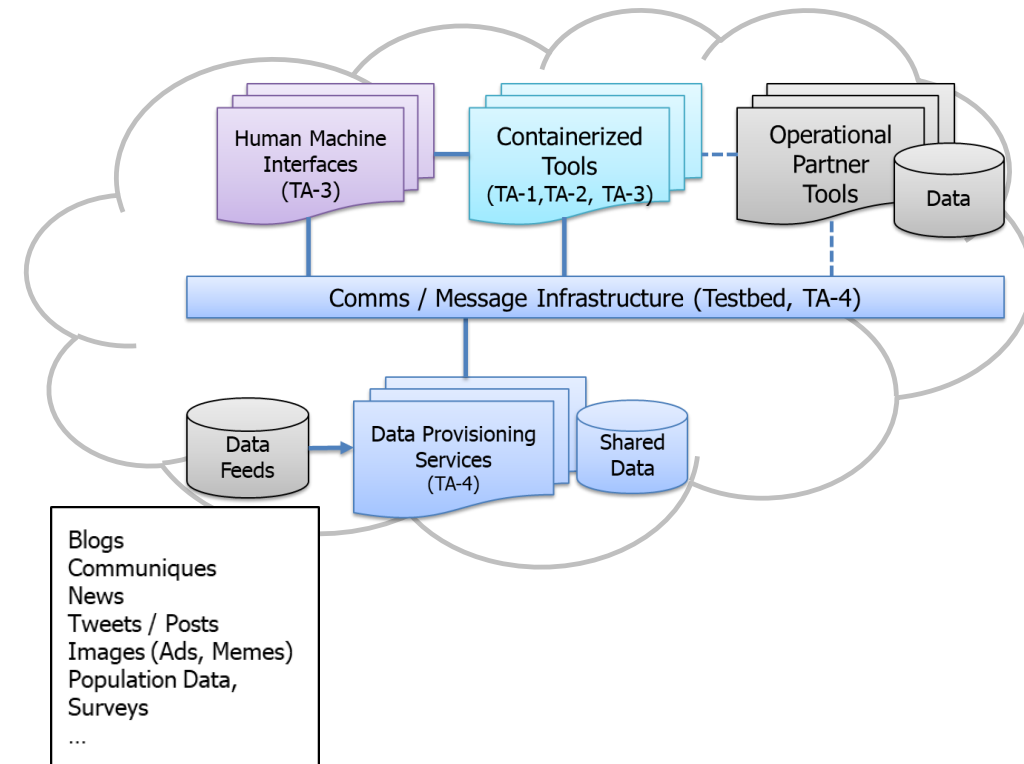
Metrics

- Provision data for 5 scenarios and 3 languages (including English)
- Support SUNet (or similar) deployment (Phase 2) & operational toolkit integration (Phase 3)

Describe

- Candidate data sources (for costing, use two realistic campaign examples)
- Testbed architecture and services for ingest, storage, analysis: utility, flexibility, and scalability
- Testbed deployment and user support (will require some personnel cleared at least to the SECRET level)
- Tool development, demonstration, and evaluation support
- Plan to work with TA1-TA3 on data requirements, API definition, testbed support

Containerized-approach and cloud-based testbed enables flexible deployment and use of INCAS tools within *existing* tool suites and systems





Evaluation (TA5) (UARC or FFRDC)

TA1: Influence Indicator Detection – T&E objectives:

1. Are messages which contain influence indicators being identified correctly?
 - Metric: f-score against gold standard corpus
2. For those messages, are influence indicators being extracted correctly?
 - Metric: f-score against gold standard corpus
 - Performers must extract agenda, concerns, and emotion indicators and additional indicators per phase

TA2: Population Response Characterization – T&E objectives:

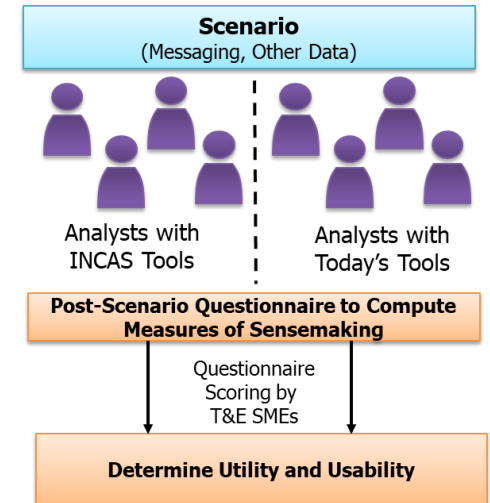
1. Are the attributes of the population segment accurate?
 - Extract demographic and psychographic attributes (see table)
 - Metric: f-score metric for users, where one can generate a gold standard
2. Are correlations among influence indicators and population segment attributes accurate?
 - Given a set of influence indicators, compare estimated attributes of responding population segment to actual attributes
 - Given a set of population segment attributes, compare estimated influence indicators they will respond to with actual indicators

TA3: Influence Campaign Modeling – T&E objectives:

1. Are the influence indicators and population segment attributes useful for sensemaking?
 - Experiments on analysts with/without INCAS tools to assess TA3 sensemaking questions (see diagram)
2. Usability of INCAS tools and calibration, utility, and usability of confidence assessment information
 - Leverage after-scenario questionnaire (e.g. System Usability Scale – usability.gov)

Program SMEs Group: Evaluation team to form a multidisciplinary group to provide knowledge of extant tools, theoretical frameworks, gaps – e.g., analysts, marketing/strategic comms, social scientists

Sensemaking Evaluation (TA3) (2 analyst cohorts, fixed time)



Use analysts or student analysts
(e.g., from USAJFKSWCS, NIU, JMITC,
Mercyhurst Univ. or UARC)



Program Metrics

TAs	T&E Objective	Metrics	Phase 1 Objectives (18 months)	Phase 2 Objectives (18 months)	Phase 3 Objectives (12 months)
Data and testbed development (TA4)	Does the program have the data and testbed for tool development and evaluation?	Scenarios	2 Historical (e.g. SocialSim Syria scenario)	2 Ongoing (e.g., China's Belt and Road Initiative)	Operationally-Relevant Scenario (with Op. Partner)
		Media	News & Social Media, English + 1 Non-English	1 Additional Non-English Language	Operationally-Relevant Data (with Op. Partner)
Influence indicator detection (TA1)	Are messages which contain influence indicators being identified correctly?	Classification accuracy	F-score > 0.8	F-score > 0.9	F-score* > 0.9
	For those messages, are influence indicators being extracted correctly?	Indicator extraction accuracy (Agenda, Concerns, Emotions)	+ 2 additional influence indicator types Average F-score > 0.7	+ 2 additional influence indicator types Average F-score > 0.8	Average F-score* > 0.8
Population response characterization (TA2)	Are the attributes of the population segment accurate?	Attribute extraction accuracy	2 demographic attributes All F-score > 0.85	+2 demographic attributes All F-score > 0.9	Demographic attributes All F-score* > 0.9
			2 psychographic attributes Average F-score > 0.7	+2 psychographic attributes Average F-score > 0.75	Average F-score* > 0.8
	Are the correlations among influence indicators and pop. segment attributes accurate?	Accuracy of estimated influence indicators and pop. segment attributes from out-of-sample data	Accuracy > 0.7	Accuracy > 0.8	Accuracy* > 0.8
Influence campaign modeling (TA3)	Are the influence indicators and population segment attributes useful for sensemaking?	Effects size of INCAS tools on sensemaking measures	Cohen's d** \geq 0.5 (medium effect size) on 2 or more measures	Cohen's d \geq 0.8 (large effect size) on 3 or more measures	Cohen's d \geq 0.8 (large effect size) on all measures*
	Usability of INCAS tools	Usability (scale 0-1)	0.7	0.8	0.8*

*During operational testing, in Phase 3, metrics will be computed/estimated against analysts/SME judgement

** Cohen's d = mean difference between two groups, divided by the pooled standard deviation



Program schedule and milestones

	Phase 1 – 18 months	Phase 2 – 18 months	Phase 3 – 12 months
	2 historical scenarios, English + 1 additional language	2 current scenarios, English + 1 additional language	Operational scenario (capstone evaluation with operational partner)
TA1 Influence Indicator Detection	Core indicator types (Agenda, Concerns, Emotion) + 2 additional indicator types	Accuracy enhancements to Phase 1 indicator types; 2 additional indicator types	Accuracy enhancements to Phase 1-2 indicator types; application to operational scenario/data
TA2 Population Response Characterization	2 demographic and 2 psychographic attributes; Segmentation and response analysis	Accuracy enhancements to Phase 1 attributes; 2 additional demographic and 2 additional psychographic attributes	Accuracy enhancements to Phase 1-2 attributes; application to operational scenario/data
TA3 Influence Campaign Modeling	Initial analyst HMI and supporting analytics; Campaign model design; Confidence assessment design and information capture.	Refine HMI and analytics based on user feedback. Confidence assessment integration & presentation.	Refine HMI based on operational scenario and user feedback
TA4 Data and Testbed Development	Data provisioning capability (text only); Testbed infrastructure/APIs, cloud deployment	Handling of images (memes, ads); Additional data sources and low-level analytics	Tool standup/integration for capstone; Addtl. data sources/low-level analytics
TA5 Evaluation	Operational stakeholders group (quarterly); Scenario development and data delivery; Evaluation		
PI Meeting	Timeline with 8 red square markers		
Quarterly Status Review	Timeline with 7 green triangle markers		
Evaluation	Timeline with 5 orange cross markers		



Additional events

- Additional Phase 1 events
 - Program Kickoff at approximately one month after program start
 - Preliminary Design Review for all TA1-TA4 performers to be held six weeks after program kickoff
 - Critical Design Review (CDR) to be held at three months after kickoff, in conjunction with the first Quarterly Status Review
- *Each phase* has several regularly occurring events that all performers are expected to participate in:
 - Evaluations
 - Two evaluation events will be held in the 9th and 16th months for both Phase 1 and Phase 2.
 - Phase 3 will feature a single, capstone evaluation 3 months prior to the end of that phase.
 - Evaluations will last up to 1 week. These will involve significant coordination across performer teams.
 - TA5 will start approx. 3 months before main effort to build out scenarios, training/test data, etc.
 - PI meetings and concurrent capability demonstrations will be held every six months
 - Status reviews will be held every three months
 - Every other review will be in conjunction with a PI meeting
 - One review per year will be held in conjunction with a site visit by the DARPA PM team to the site of each team's prime contractor
 - Remaining reviews will be held virtually via teleconference or videoconference (e.g. Zoom.gov, Microsoft Teams, etc.)
 - Short informal monthly status meetings will be held with each performer team via teleconference or videoconference



Program deliverables

- Each TA1-TA4 team is expected to deliver the following at least one month before each evaluation event and at the end of each program phase:
 - source code and documentation, build scripts
 - containerized executable code (in Phases 2-3)
 - documentation to install, run, and operate the software (as applicable)
- Each TA1-TA4 team is expected to deliver
 - Program Kickoff Brief, a Preliminary Design Review brief at six weeks after kickoff
 - a Critical Design Review brief at three months after kickoff,
 - a Final Report (MS Word) at 47 months (or the end of the contract, whichever comes first).
- Each TA1-TA4 team is expected to deliver technical and financial status reports every month along with briefings for Quarterly Status Reviews and PI meetings.



Additional Responsibilities

- Privacy Protection
 - Appropriate controls with respect to collection of data on US persons, privacy, personally identifiable information (PII), etc.
 - TA4 proposers must detail how they plan to implement these controls, with specific attention to identifying and removing PII when necessary
 - TA4 performers will be responsible for providing relevant data policies and training to all performers
- Human Subjects Research Controls (where applicable)
- Associate Contractor Agreements



Proposal Information

- Performers may submit TA1-TA4 proposals.
- Each proposal may only address one TA.
- Separate proposals for each TA are required if proposing to multiple TAs.
- Proposers addressing multiple TAs must describe expected research synergies and specific efficiencies and savings to the Government that would result if that proposer was selected for multiple TAs.
- If a proposer proposes to both TA4 and either TA1 or TA2, that proposer must address in the proposal how TA4, as data provider and testbed developer, will ensure a “level playing field” to all other TA1 or TA2 performers.
- Teaming is encouraged but not required. Consider multi-disciplinary perspectives (computer science/AI, social science, marketing, etc.).
- Multiple awards each for TA1 and TA2. Single award for TA3. Single award for TA4.



www.darpa.mil