# Artificial Intelligence Lab Research Bootcamp: Conducting Computational IS Research

Sagar Samtani, Ph.D.

Assistant Professor, Grant Thornton Scholar, and CACR Fellow

Kelley School of Business, Indiana University

**Acknowledgements:** Drs. Hsinchun Chen, Jay F. Nunamaker, Jr., Sue Brown, Mark Patton, Byron Marshall, and hours of conversation with Drs. Hongyi Zhu, Yidong Chai, and Balaji Padmanabhan



1

## Outline

- **About Me**
- **Bootcamp Purpose and Materials**
- **Bootcamp Modules:**
  - **Module 1:** Background and Importance of Research?
  - **Module 2:** Conducting Research in the AI Lab
  - **Module 3:** Presentation and Professional Progression
- **Conclusion**



2

# About Me

- Assistant Professor and Grant Thornton Scholar in the Department of Operations and Decision Technologies at Kelley School of Business at Indiana University.
  - Assistant Professor at University of South Florida from 2018 – 2020.
- Ph.D. (SFS and AI Lab alum; 2018), MS (2014), BS (2013) at UArizona MIS.
- **Current Research Interests:**
  - **Domain** – cybersecurity → cyber threat intelligence; AI for cybersecurity; scientific cyberinfrastructure; vulnerability management; Dark Web analytics; social media analysis
  - **Methods** – Artificial Intelligence → deep learning (GCN, GAN, DSSM, interpretability), network science, diachronic linguistics, explainable AI

3

3

# Bootcamp Purpose

- **Bootcamp objective** → fast-track you to conducting computational information systems (IS) research:
  - Introduction to what research is and is NOT.
  - Summarize value of research to the institution, you, and society.
  - Walkthrough of academic research templates.
    - Literature review, research design, novelty, evaluation, etc.
  - Academic research considerations (e.g., presentation, papers, etc.).

4

4

# Disclaimers!

- **Disclaimer 1:** I do not know everything.
- **Disclaimer 2:** The views and opinions in these slides reflect mine only and may change or evolve.
- **Disclaimer 3:** When presented live, these slides are supplemented by numerous examples of the concepts described. However, the slides are sufficiently detailed such that readers can extract the main points without a live presentation.
- **Disclaimer 4:** Just reading these slides is not enough; experience is king.

5

5

# Before we start…

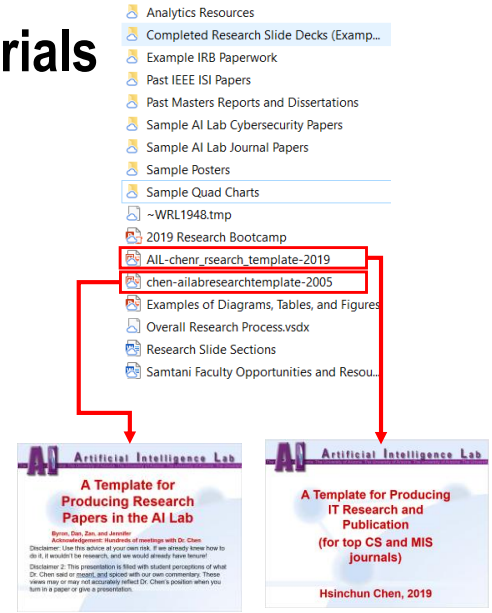
- Guiding questions for you:
  - Who is your advisor?
  - What year are you?
  - What is your area (or interests) of academic research?
  - **What is your vision for yourself in the next 1, 2, 5, and 10 years? Where do you want to be?**
- Carefully think about these throughout the presentation.
- Link what is talked about to yourself.

6

6

# Bootcamp's Provided Materials

- This bootcamp has a "care package" Zip folder with:
  - This set of slides
  - Past journal papers
  - Past master's papers
  - Example slide decks, quad charts, posters
  - Journal and conference resource list
  - Dr. Chen's research templates (These slides are not meant to replace those!)
  - General research process
  - Analytic tools inventory




7

7

# Bootcamp Modules

- **Module 1: Background and Importance of Research**
- **Module 2: Conducting Research in the AI Lab**
- **Module 3: Writing Skills, Presentation Tips, and Professional Progression**
- **Conclusion and Wrap-up**

8

8

# M1: What is Research?

- The term *research* is often loosely used in general society.
- Academic research **IS NOT JUST:**
  - Looking up information (i.e., information discovery)
  - Building tools and systems
  - Mere technical improvement of systems
  - Regurgitation or organization of already known information
- However, all of the above can be **part of the research process**.

9

9

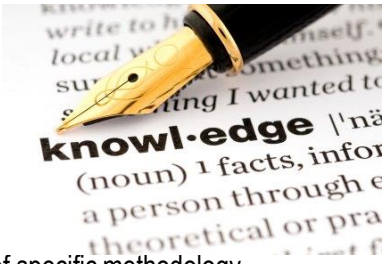
# M1: What is Research?

- Academic research is the systematic process of **creating knowledge.**
- Academic research **observes, describes, predicts, and explains** the world.
- Academic research should be **repeatable and observable.**
- We want to systematically understand specific **phenomena of interest to create knowledge**.

10

10

# M1: What is Research?




- **Academic research involves:**
  - Clear articulation of a problem and its context
  - Specific, systematic plan of addressing the problem
  - The testing of a specific theory and/or utilization/development of specific methodology
  - Collection, analysis, and interpretation of data
  - Implications and impact of research
  - …
- **Importance of Research:**
  - **To the institution:** brand recognition, better students, research money, commercialization (e.g., Google, COPLINK), societal impact, innovative course development, etc.
  - **To you:** security is evolving; need to stay updated → makes you more valuable
    - Rare amongst SFS programs! Most focus on classes only.

11

11

# M1: Importance of Research

- **Research develops critical skills in an evolving cybersecurity landscape:**
  - **Value (why > what):**
    - What **value are YOU offering**? (e.g., Literature, practical, theoretical, etc.)
    - What value do certain tools offer? (e.g., Nessus vs Burp Suite, Zmap vs Nmap, etc.)
    - What value do certain techniques/methods have? (e.g., Random forest vs neural nets)
  - **Decomposition:** breaking down a problem into sub-problems.
  - **Critical thinking:** developing your own critical understanding of an area.
  - **Vision:** what do you want the paper to look like? How does it fit into your career? Your objectives? Etc.
  - Etc.

12

12

# Outline of Bootcamp

- Module 1: Background and Importance of Research
- **Module 2: Conducting Research in the AI Lab**
- Module 3: Presentation and Progression
- Conclusion and Wrap-up

13

13

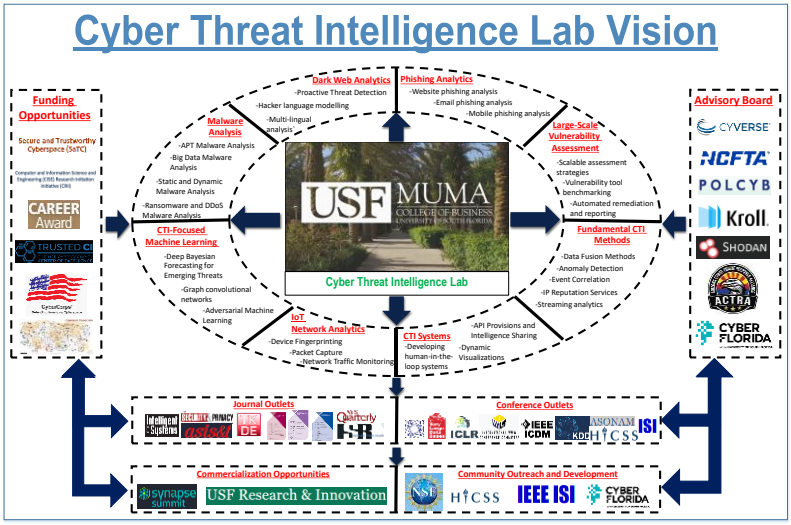
# M2: What is MIS?

- **Definition:** MIS is the academic discipline that studies the creation, adoption, and use/impact of information technology for business and/or societal applications.
  - **Creation** → design systems/algorithms; **Father:** J. Nunamaker; **Schools:** UA, USF, NYU
  - **Adoption** → behavioral (theory/survey/experiment); **Father:** G. Davis; **Schools:** UMN, GSU
  - **Use** → Economics/econometrics; **Father:** A. Whinston; **Schools:** UT Austin, ASU, UMD
- Business/societal applications → cybersecurity, health, e-commerce, etc.
- **Differences between IS and CS:**
  - IS → Information centric; data driven; application driven
  - CS → foundations of computing (e.g., computing theory)

14

14

# M2: How a Lab Operates




- **Key Benefits of a Lab:**
  - Build teams around major real-world issues
  - Infrastructure → students, software/hardware, personnel
  - Develop reputation → attract students, funding, etc.
  - Community development
  - Course/education integration
  - **AI Lab** → $50M+ funding; 35 Ph.D. students; COPLINK system; world-class director, etc.

15

15

# M2: AI Lab Template

- **AI Lab Template:**
  - Topic selection
  - Title
  - Abstract
  - Introduction
  - Literature Review (Domain+ Method)
  - Research Gaps and Questions
  - Research Testbed and Design
  - Results and Discussion
  - Conclusion and Future Direction

- **AI Lab Template Operationalized:**
  1. Topic selection
  2. Literature Review (Domain)
  3. Literature Review (Method)
  4. Research Gaps and Questions
  5. Research Design
  6. Results and Discussion
  7. Conclusion and Future Directions
  8. Introduction
  9. Abstract
  10. Title

**Iterative, ongoing Process; do not be afraid to fail!**

16

16

# M2: AI Lab Template – Selecting a Topic

- **According to Dr. Chen:**
  - "Research topic needs to be new and interesting" → "Avoid old and well-studied topic"
  - "Research could be technique/algorithm driven or application driven"
  - "Read a lot. Understand current trends and directions"
  - "Use well grounded methodologies"
  - "Compare with existing techniques/approaches with data sets"
- Choose based on NSF, media, or stakeholder (e.g., CyVerse, LEO, NCFTA, POLCYB, etc.) needs and interests.
  - Relate back to your bigger vision. Unlikely you can make big impact with one paper.
- If no one is researching a topic, ask why before pursuing it.
  - Lack of technology, not an interesting problem, not a real issue, etc.

17

17

# M2: AI Lab Template – Literature Review

- **IS → information centric; data-driven; application driven**
  - Develop a novel method to solve an important problem within a domain (application)
  - Need to show you know what the field looks like; grounding and justification for your approach.

- **Domain Specific Literature Review:**
  - Key definitions and background
  - Who (major groups/scholars) has done what?
  - What does their data look like (e.g., characteristics)?
  - What have their approaches been (e.g., manual/automatic)?

- **Method Specific Literature Review:**
  - Where does the method originate from (i.e., what class of methods)?
  - Why this method? What are the other options?
  - How does the method operate? (math, diagram)
  - Who uses the method? For what application?

**Ongoing, iterative process. Always monitor, re-read, and update your review.**

18

18

# M2: AI Lab Template – (Domain) Literature Review

- Places to get literature → venues that emphasize domain + method
  - **Seeds:** Survey paper (e.g., ACM CSUR, IEEE C&T); NSF (SaTC PIs); keywords
  - **Security Journals:** ACM TOPS, ACM DTRAP, TDSC, TIFS, DI, C&S, IoTJ
  - **Security conferences:** ISI, CAMLIS, IFIP, SKM, WEBEIS, USENIX (ScAINet Workshop), NDSS, IEEE S&P (DLS Workshop), ACM CCS (AISec workshop)
  - **Application oriented CS (conferences):** WWW, ASONAM (FOSINT-SI Workshop), KDD, ICDM (DL-CTI Workshop), VLDB, AAAI, ACL
  - **Other:** criminology venues
- Aggregate papers (use Mendeley + Google Scholar) from past 3-5 years:
  - Find paper → extract references→ repeat till completed
  - **Paper quality:** venue, authors, quick read (abstract, intro, conc., figures/tables)

19

19

# M2: AI Lab Template – (Domain) Literature Review

- How to review → Extract each paper's key info into a table (taxonomy).
  - Year (recent first), authors, venue, dataset (size, coverage, language, etc.), method, selected results, other key details
    - Common mistake → too many words in each cell
- **Structure in slides:**
  - Key definitions and background of the domain
  - Table summary of recent and relevant literature
  - Key observations (answers to key questions regard data, coverage, approaches, etc.)
  - Transition to method or another related area of domain

20

20

# M2: AI Lab Template – (Method) Literature Review

- Select a method based on prior (domain) literature and/or data characteristics.
- Places to get literature → venues that emphasize method
  - **Seeds:** Survey paper (e.g., ACM CSUR); NSF (SaTC PIs); keywords
  - **CS Conferences: NeurIPS, ICLR,** ICDM, ICML, ASONAM, KDD
  - **CS Journals:** IEEE IS, IEEE TKDE, IEEE PAMI, IEEE TNNLS, ACM TOIS, IEEE TEM, ACM TIST
- Aggregate papers (use Mendeley + Google Scholar) from past 3-5 years:
  - Find paper → extract references→ repeat till completed
  - **Paper quality:** venue, authors, quick read (abstract, intro, conc., figures/tables)





# M2: AI Lab Template – (Method) Literature Review

- How to review → Extract each paper's key info into a table (taxonomy).
  - Year (recent first), authors, venue, dataset, method, selected results, evaluation procedures, other key details
    - Common mistake → too many words in each cell
- **Structure in slides:**
  1. Justification for the method
  2. Background of the method (where did it originate → class of methods)
  3. Key definitions and operations (math, diagram, key steps)
  4. Table summary of recent and relevant literature
  5. Key observations (limitations)
  6. Transition to another method or research gaps and questions

# M2: AI Lab Template – Research Gaps and Questions

- **Example domain specific research gaps:**
  - Methods are manual instead of automated (lacks scalability)
  - Have not examined a particular phenomena (e.g., identifying key hackers)
  - Did not account for specific data characteristics (e.g., missing features)
  - Etc.

- **Methodological research gaps:**
  - Did not capture specific features, representations, or encodings
  - Past methods were not scalable or manual
  - Past method was only supervised or unsupervised
  - Past method did not integrate specific functionality (e.g., interpretability)
  - Etc.

23

23

# M2: AI Lab Template – Research Design

- **Four (minimum) components to a research design:**
  1. Data collection (i.e., research testbed)
  2. Method/system/algorithm (i.e., core novelty)
  3. Evaluation
  4. Case Study

- Show one professional diagram with all interlinking components.
  - Very useful for external presentations.

- Each step in your research design must be justified based on:
  - Prior literature (including relevant theory, if applicable).
  - Data characteristics
  - Objective of your study

24

24

# M2: AI Lab Template – Research Design (Data Collection)

- From Dr. Chen:
  - Use research testbed to validate designs and approaches.
  - What data sets will be used in the experiment or evaluation?
    - Construct gold-standard dataset
  - Testbed should be interesting, relevant, and significant. Size and scale matter.

- How to understand your data:
  - How populated?
  - How many duplicates?
  - Key metadata (number, date range, categories, other descriptive statistics)
  - Key features within text (e.g., # of keywords, etc.)
  - Pre-processing to clean data (e.g., stemming, lemmatization, stopword removal, etc.)

25

25

# M2: AI Lab Template – Research Design (Method)

- Method → algorithm, system, framework, etc. → how you are solving the problem.

- All novelties should be inspired by the key characteristics of the application (e.g., data, tasks, requirements, processes, theories, etc.).

- **Four types of novelty (or a mix):**
  - **Application** → using existing method on new dataset
    - Weakest; only works if application is very new or past approaches have been manual/old algorithms
  - **System** → multiple, disparate components (e.g., algorithms) linked together
  - **Representation** → capturing additional data features (must be clearly justified)
  - **Algorithmic** → adjust internal components of the algorithm (new mechanism, unsupervised to supervised, convolutional operations, filters, custom SVM kernel, etc.)

26

26

# M2: AI Lab Template – Research Design (Method)

- **How to present in slides:**
  1. Briefly summarize past approach(es) → how they work, why they are not suitable
  2. Provide diagram of your proposed method → highlight in red your novelty
  3. Justify each component (based on literature or data characteristics)
  4. Illustrate key math and algorithmic detail (e.g., pseudocode, dedicated sub-sections)

- Be crystal clear on the key technical differences and their resultant domain benefits of your approaches.
  - Lists and tables are very valuable in this regard.

27

27

# M2: AI Lab Template – Research Design (Evaluation)

- **Objective:** Evaluate each component you are claiming contribution to.
  - **Justification:** Need to show your method outperforms the best in the field.

- **Five major components to a thorough, convincing evaluation:**
  1. **Dataset:** Ground-truth datasets model training, testing, and benchmarking
  2. **Model Training and Testing:** Training and testing your proposed model
  3. **Model Benchmarking:** Systematically compare the proposed model against benchmarks
  4. **Post-hoc (i.e., post-model training) analysis:** sensitivity of the model to various settings
  5. **Interpretation and Insights (i.e., Technical Case Study):** Demonstrate the value of your work!
- Quantity, depth, and coverage of these components will vary based on the study.

28

28

# M2: AI Lab Template – Research Design (Evaluation)

| Component | Key Aspects | Description | Example(s) |
|---|---|---|---|
| Dataset | Ground-truth dataset construction | Labelled dataset used for model training and testing representative of the phenomena of interest | Complete dataset fully labelled by experts |
| | Train | Portion of data that is used to train the algorithm(s) | Randomly selected 80% of the ground truth dataset |
| | Development (i.e., tuning) | Portion of data that is used to tune the algorithm(s) | Randomly selected 10% of the ground truth dataset |
| | Testing | Portion of data that is used to test and evaluate algorithm performance | Randomly selected 10% of the ground truth dataset |
| Model Training and Testing | Hyperparameter selection | Selecting values to control the learning process | Grid-search, pre-optimized or trained model |
| | Training Strategy | How the proposed model is trained and the model parameters learned | 10-fold cross validation, hold-out, pre-trained model, training strategy based on tests for overfitting and underfitting |
| Model Performance Benchmarking | Performance Metric Selection | Metrics to evaluate the performance | Accuracy, precision, recall, F1, NDCG, MAP, MRR, homogeneity, NMI |
| | Evaluation against non-DL models | Proposed DL model vs non-DL-based models | Naïve Bayes, SVM, Decision Tree, k-NN |
| | Evaluation against DL models | Proposed DL model vs prevailing DL-based models | CNN, LSTM, GRU, RNN, ANN |
| Post-hoc (i.e., post-model training) evaluation | Sensitivity or Ablation Analysis | Internal analysis of DL model to interpret how model components contribute to overall performance | # of layers, activation functions, varying model components, counterfactual analysis |
| | Convergence speed | How quickly the model converges | Speed, computational complexity |
| | Model stability | How stable the model is in training, comparison, etc. | Validation loss, thresholding, statistical significance |
| Interpretation and Insights (Technical case study) | Examples of outperformance | Identifying where the proposed algorithm outperformed baselines | Identify 1-2 instances within the ground-truth dataset that were correctly identified by the proposed method, but missed by the best competing benchmark |
| | Apply proposed DL on unseen data | - | Applying a transfer learning framework to categorize all hacker exploits in forums |



29

# M2: AI Lab Template – Research Design (Evaluation – Dataset, Training, and Testing)

- Construct representative, and comprehensive gold-standard dataset.
  - Inter-coder reliability (cohen's kappa) when gold-standard does not exist
  - Publicly accessible gold-standard datasets (to show cross-domain generalizability)
- In supervised tasks, gold-standard datasets are used to construct:
  - **Training:** Portion of data that is used to train the algorithm(s)
  - **Development (i.e., tuning):** Portion of data that is used to tune the algorithm(s)
  - **Testing:** Portion of data that is used to test the algorithm(s)
- Model parameters are learned during training and hyperparameters are tuned on the development set.



30

# M2: AI Lab Template – Research Design (Evaluation – Model Performance and Benchmarking)

- **Use a table to present three sets of evaluations (for DL-based studies):**
  - **Set 1:** Evaluation against standard ML algorithms (NB, SVM, RF, DT, ANN)
  - **Set 2:** Evaluation against DL algorithms (e.g., CNN-based approaches, LSTM-based approaches, etc.)
  - **Set 3:** Sensitivity analysis (i.e., internal or ablation analysis) (e.g., # of layers, adjustments to inputs, etc.)

- **Performance Metrics (perform paired t-tests or statistical tests for all):**
  - **Classification:** accuracy, precision, recall, F1, ROC, AUC
  - **Clustering:** homogeneity, completeness, NMI, Rand Index, v-measure, Calinski Harabaz, Silhouette
  - **Info. Retrieval:** NDCG@K, MAP, MRR, P@K, precision curves
  - **Others:** perplexity, block/edit distance, BLEU, reconstruction

- **Post-hoc model evaluations (sometimes requested by reviewers):** qualitative, convergence analysis, complexity analysis, computational time, etc.

31

31

# M2: AI Lab Template – Research Design (Technical Case Study)

- In technical (computational, data-driven) IS research, we need to show the value of the proposed method with a case study.

- **Purpose of Case Study:**
  - Illustrate examples of where your method outperformed benchmark methods
    - What does the proposed method discover that previous ones missed?
  - Apply your method to unseen data. What can you discover?
    - Helps illustrate potential value of the approach. Figures, charts, and visualizations are exciting!
  - Demonstrate with user evaluations usefulness and value (MISQ/ISR)

32

32

# M2: AI Lab Template – Research Design (Method – Sample Tools)

- It is important to innovate on the methodology, but no need to re-invent the wheel in terms of the process.
- There are many tools that may have already implemented what you are looking for!
- Following two slides summarize popular data, analytics, and presentation tools that can help develop methodologies:
  - Important to use whichever tool is needed to execute your research.
  - Automate as much of the work as you can!
  - Look for open source GitHub code releases (e.g., paperswithcode.com)

33

33

# M2: AI Lab Template – Research Design (Method – Sample Tools)

| Category | Task | Tool/Package Name(s) | Description | Documentation |
|---|---|---|---|---|
| Data | Collection | Scrapy | Python package for incremental web crawlers | https://scrapy.org/ |
| | | JSON | Python package for parsing JSON data from APIs | https://docs.python.org/3/library/json.html |
| | | BeautifulSoup | Python package for general web crawling | https://pypi.org/project/beautifulsoup4/ |
| | | Offline Explorer | GUI-based web crawling software | https://offline-explorer.en.softonic.com/ |
| | Storage | MySQL | Relational database | https://www.mysql.com/ |
| | | MongoDB | NoSQL database | https://www.mongodb.com/ |
| | | Pickle | Python package for storing ML/DL models | https://docs.python.org/3/library/pickle.html |
| | Summary statistics | Data Analysis Baseline Library | Common ML pre-processing tasks | https://amueller.github.io/dabl/dev/ |
| | | Pandas | Formatting and structure data inputs from varying data sources | https://pandas.pydata.org/ |
| | | SideTable | Advanced data-wrangling with Python | https://pbpython.com/sidetable.html |

34

34

# M2: AI Lab Template – Research Design (Method – Sample Analytics and Presentation Tools)

| Category | Task | Tool/Package Name(s) | Brief Description | Documentation |
|---|---|---|---|---|
| Analytics | General ML | Scikit-learn | Python package for basic ML algorithm implementation | https://scikit-learn.org/stable/ |
| | | RapidMiner, WEKA | GUI-based, general purpose ML toolkits | https://rapidminer.com/; https://www.cs.waikato.ac.nz/ml/weka/ |
| | Text Analytics | NLTK | Python package for symbolic and statistical NLP | https://www.nltk.org/ |
| | | Spacy | Industrial strength, large-scale information extraction and NLP | https://spacy.io/ |
| | | Stanza | Python package from Stanford for multi-lingual analysis | https://stanfordnlp.github.io/stanza/ |
| | Deep Learning | Pytorch | Advanced Python package for customizable deep learning | https://pytorch.org/ |
| | | Keras | Basic package with standard DL algorithms | https://keras.io/ |
| | | fastai | Various tools and resources for DL | https://www.fast.ai/ |
| | Network Science | Networkx | Python package for basic network science tasks | https://networkx.github.io/ |
| | | igraph | Advanced python package for extensive network science | https://igraph.org/python/ |
| | | stellargraph | Graph embedding package | https://github.com/stellargraph/stellargraph |
| | | Gephi | GUI-based tool for network science | https://gephi.org/ |
| Presentation | Visualization | Tableau | GUI-based tool for visualizations | https://www.tableau.com/ |
| | | Matplotlib | Basic Python-based visualization package | https://matplotlib.org/ |
| | | Plotly | Advanced Python-based visualization package for ML/DL | https://plotly.com/ |
| | Web front-end | Streamlit | Python package for rapid prototyping of DL/ML-based systems | https://www.streamlit.io/ |
| | | Django | Python-based web application technologies | https://www.djangoproject.com/ |

35

35

# M2: AI Lab Template – Research Design (Results and Discussion)

- Results and Discussion should have two levels of discussion:
  1. Presentation and Discussion of Technical Results
  2. Presentation and Discussion of "Non-Technical" (domain-relevant) Results

- **Technical results:**
  - Present your tables and graphs of results.
  - Discuss what outperformed. For what reasons?
  - Illustrate 1-2 meaningful examples where your proposed method detected

- **"Non-Technical" domain-relevant results:**
  - Apply your novel method on unseen data. Need to attain, actionable, prescriptive, "cool" results.
  - What comes up? What did you detect? Be thoughtful, top-down, and systematic of the results.
  - Don't overstate the contributions. Don't speculate without evidence.

36

36

# M2: Introduction, Conclusion, Abstract, Title

- Will have some duplication with abstract and introduction.
- Re-state the problem → importance, scale, etc. (2-3 sentences)
- Summarize prior approaches (e.g., research gaps) (2-3 sentences)
- Summarize your approach and contributions (3-4 sentences)
- Future directions → significant extensions that can lead to more papers.
  - **Common mistake:** future directions are too simple (e.g., more experiments)

37

37

# M2: Introduction, Conclusion, Abstract, Title

- Introduce and motivate the topic (sell your work):
  - Stats, figures, etc. → literature here should be drawn from NSF, newspapers, etc.
  - Illustrate the data with a screenshot → boxes, arrows, and brief descriptions.
- Have a clear, crisp, and unambiguous problem specification.
- Summarize your approach, with focus on your novelty.
  - Methodological and domain contributions. Present as clear, concise, unambiguous bullet points.
- Common mistakes:
  - Not showing the scope, scale, size, importance of the problem → is this a real problem?
  - Not showing characteristics of the data → what does the data look like? **A carefully constructed screenshot can go a long way!**
  - Not being clear what the contribution is → bullet points of the key novelties can make it very clear

38

38

# M2: Introduction, Conclusion, Abstract, Title

- **Abstract (from Dr. Chen):**
  - Most important part of a paper – the first impression!
  - Abstract should reflect the entire paper.
  - 200-300 words in one paragraph.
  - 2-3 sentences to summarize problem motivation.
  - 2-3 sentences to describe proposed method or algorithm.
  - 2-3 sentences to summarize evaluation method.
  - 3-4 sentences to summarize key findings.
  - Write abstract after finishing the entire paper. Select key sentences from paper.
- **Title (from Dr. Chen):**
  - 8 words or less.
  - Develop a title after finishing the paper.
  - Title needs to reflect the essence of the research.
  - Don't use cute title, e.g., "To aggregate or not to aggregate"
  - Use project/system acronym with clear relevant meaning, e.g., COPLINK, BioPortal; not ALOHA.

39

39

# Outline of Bootcamp

- Module 1: Background and Importance of Research
- Module 2: Conducting Research in the AI Lab
- **Module 3: Presentation and Progression**
- Conclusion and Wrap-up

40

40

# M3: Why Slides First?

- Dr. Chen's "secret sauce" for conducting research.
  - Preferred method of communicating research progress.
- Promotes:
  - Good, concise writing
  - Good structure
  - Easy transfer over paper
  - Easy presentation at conferences and other external venues
- Becoming a preferred method of professional communication.
  - Ability to present work at varying levels of granularity.

41

41

# M3: Overarching Slide Principles

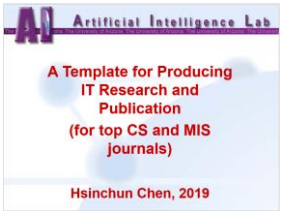
- Follow the template.
- Slide numbers on every slide.
- Simple format/background (red, white, blue, black).
- Every table and figure has a caption and is referred to in the text.
- Consistent format (e.g., tables, figures, references).
- Max four, two line main bullet points per slide.
- No typos or grammatical errors! Hire professional copy-editors.
- Transition and flow is critical.
- Following these principles allows Dr. Chen to focus on the content of your message and provide helpful research comments.
  - Avoid comments about structure, writing, etc.

42

42

# M3: Transitioning Slides to Paper

- Dr. Chen's templates provide excellent advice on paper structure.
  - Provides details (e.g., length, content, etc.) for each section.




- My process:
  - Open blank document → put in paper structure.
    - Introduction, literature review, research gaps and questions…
  - Copy, paste, and adapt slide contents in.
  - Smooth out writing by adding transitions between major sections.
  - Read out loud to catch errors.
  - Have peers review and provide feedback.
  - Then copy into the publication format (e.g., IEEE).
    - Closely follow their requirements.

43

43

# M3: Transitioning Slides to Quad Chart

- Technical document used to briefly describe your work.
  - Very colorful, illustrative, highlights of your overall research
- Divided into four quadrants.
  - Research focus, approach, results, impact
- Very valuable at SFS job fair.
  - Provided some examples of your work, tools, etc.

44

44

# M3: Transitioning Slides to Poster

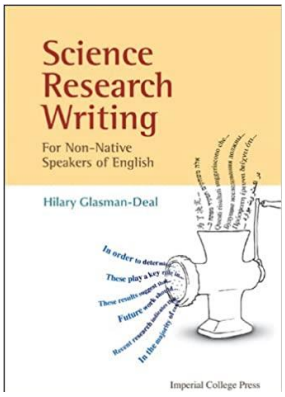
- Sometimes you will be asked to present your work as a poster at conferences.
  - Examples provided
- From NYU (http://guides.nyu.edu/posters):
  - Important information should be readable from about 10 feet away
  - Title is short and draws interest
  - Word count of about 300 to 800 words
  - Text is clear and to the point
  - Use of bullets, numbering, and headlines make it easy to read
  - Effective use of graphics, color and fonts
  - Consistent and clean layout
  - Includes acknowledgments, your name and institutional affiliation

45

45

# M3: Writing Tips and Tricks

- Omit needless words.
- Write everyday.
- Read your writing out loud.
- Do I really need this word/sentence/paragraph here or at all?
- Each sentence/paragraph needs to convey one thought.
- Read good, systematic writing (e.g., past AI Lab papers, WSJ/NYT/WP).
- Get feedback early and often.



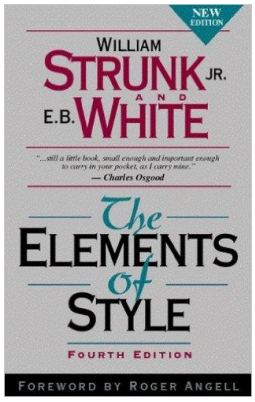




**Figure Credits: Amazon**

46

46

# M3: Research Considerations

- Humans by nature **rationalize** behaviors and/or effects and want to **minimize time and effort.**
- Very helpful from evolutionary standpoint, dangerous in research.
  - Leads to skewed/inaccurate perceptions of results.
- Want to conduct and deliver research **objectively**. Do not be afraid to pivot or take criticism (on your research and writing).
  - "Ego is the enemy," "Good is the enemy of great"
- Be aware of biases and how they may affect your work.
  - Hindsight bias, primacy/recency effect, confirmation bias, etc.

47

47

# M3: Research Considerations

- Executing research will help you "learn how to learn."
  - Excellent skill, especially for a rapidly evolving technological field.
- However, it requires:
  - Very positive attitude (not everything is going to work out!)
  - Strong work-ethic
  - Discipline and self-motivation
- Your advisor can provide advice and feedback.
  - They will not do the work for you
  - In many cases, you will have to teach them!

48

48

# M3: Research Considerations – Managing Meetings

- Regular meetings with your advisor are critical.
- AI Lab, SFS, and other research meetings with your faculty are synchronization meetings, not "brainstorming" or overly detailed technical meetings.
  - Make sure you are on track; faculty are very busy!
- Prepare your update the night before.
  - What you have been working on since last update.
  - Provide any intermediate updates to let your advisor know your status.
  - Any specific questions that you cannot figure out at all.
  - What you intend to do next.
  - What your timeline is moving forward.
  - Stay focused when delivering your update.
  - Don't talk about classes or anything unrelated to research/lab work

49

49

# M3: Research Considerations – Maintaining Operational Research Productivity

- Documenting your research is essential to maintaining good research progress.
- Common mechanisms include:
  - **IDE's and Package Management:** PyCharm, Jupyter, Anaconda Navigator
  - **Code repositories:** GitHub, Stack Overflow
  - **Communication Software:** Slack, Zoom, Skype, Teams, Outlook
  - **Citation Management:** Mendeley (with plugins), Google Scholar
  - **Note Management and Collaboration:** Confluence, Notability, Evernote
- Keeping these up to date can help you quickly develop a suite of resources to rapidly advance you research, as well as help onboard new members quickly!

50

50

# Outline of Bootcamp

- Module 1: Background and Importance of Research
- Module 2: Conducting Research in the AI Lab
- Module 3: Progression and Presentation
- **Conclusion and Wrap-up**

51

51

# Conclusion and Wrap-Up

- Research is the systematic process of creating knowledge.
- **Bootcamp objective** → fast-track you to conducting research:
  - Introduction to what academic research IS and IS NOT.
  - Summarize value of research to the institution, you, and society.
  - Walkthrough of academic research templates.
    - Literature review, research design, novelty, evaluation, etc.
  - Academic research considerations (e.g., presentation, papers, etc.).

52

52