IU DSAIL Research Bootcamp:
Conducting Computational IS Research

Sagar Samtani, Ph.D.

Assistant Professor, Arthur M. Weimer Fellow

Founding Director, Data Science and AI Lab

Kelley School of Business, Indiana University

Acknowledgements: Drs. Hsinchun Chen, Jay F. Nunamaker, Jr., Sue Brown, Mark Patton, Byron Marshall, Weifeng Li, Reza Ebrahimi, and (many) hours of conversation with Drs. Hongyi Zhu, Yidong Chai, and Balaji Padmanabhan

Outline

- **Bootcamp Purpose and Materials**
- **Bootcamp Modules:**
  - **Module 1:** Background and Importance of Research?
  - **Module 2:** Conducting Research in DSAIL
  - **Module 3:** Presentation and Professional Progression
- **Conclusion**

Bootcamp Purpose

- **Bootcamp objective** → fast-track you to conducting computational information systems (IS) research:
  - Introduction to what (computational IS) 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.).

Disclaimers!

- **Disclaimer 1:** I do not know everything. I am still learning.
- **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.
- **Disclaimer 5:** These slides cover more beginner material (e.g., getting started); activities such as IS positioning, managing revisions, positioning for grants are not covered here.

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.

Bootcamp Modules

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

M1: What is Research?

- The term *research* is often loosely used in general society.

- 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**.

M1: What is Research?

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

M1: What is Research?

- **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 firms:** Top tech firms (e.g., Google, Meta, etc.) have dedicated research teams!
  - **To you:** technology is evolving; need to stay updated → makes you more valuable

M1: Importance of Research

- **Research develops critical skills in an evolving technology 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.

Outline of Bootcamp

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

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:** UArizona, 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)

M2: How a Lab Operates

- **Key Benefits of a Lab:**
  - Build teams around major real-world issues → security, health, collaboration, etc.
  - Infrastructure → students, software/hardware, personnel, etc.
  - Develop reputation → attract students, funding, etc.
  - Corporate partnerships → Microsoft, FireEye, CrowdStrike, BAH, etc.
  - Government/non-profit partnerships → NSA, DHS, MITRE, NCFTA, POLCYB, etc.
  - Community development → workshops, conferences, case competitions, etc.
  - Course/education integration → MS in cybersecurity, SFS, etc.
  - …

CREATING HIGH-VALUE REAL-WORLD IMPACT THROUGH SYSTEMATIC PROGRAMS OF RESEARCH[1]

Jay F. Nunamaker
Eller College of Management, University of Arizona, McClelland Hall 430, P.O. Box 210108,
Tucson, AZ 85721-0108 U.S.A. {jnunamaker@cmi.arizona.edu}

Example (Non-Exhaustive) Labs in the IS Discipline

- **CMI:** UArizona; Director, J. Nunamaker; 100+ Ph.D. grads; 1984 – present; GroupSystems; GDSS, Avatar, DHS BORDERS, **100M+ funding**, 250+ journal pubs.
- **AIL:** UArizona, Director, H. Chen; 35 Ph.D. grads (six ongoing); 1989 – present; digital libraries, COPLINK, Dark Web, AZSecure, **60M+ funding**, 300+ journal pubs.
- **MIT CISR:** Coordinator: W. Orlikowski; 1974 – present; field-based research, IT consulting/management, 250+ journal pubs.
- **CDIN:** GSU; Director: A. Rai; process/digital innovation, governance; 16+ Ph.D. grads; 2004 – present; 160+ journal pubs.
- **HAL:** U. Notre Dame; Directors, A. Abbasi and K. Kelley; 2021 – present; psychometric NLP, predictive analytics; **~$550K+ funding (UND share; Abbasi);** 50+ journal pubs. (Abbasi).
- **DAL:** U. Delaware; Director, X. Fang; data analytics, AI; 2 Ph.D. grads (five ongoing); 20+ pubs.
- **BITL:** BU; Director; DK Lee; text mining; 1 Ph.D. grads (four ongoing); 10+ journal pubs
- **DSAIL:** IU; Director, S. Samtani; AI for cyber, MHI; five ongoing Ph.D.; 2022 – present; **~$5M+ funding (IU share);** 30+ journal pubs.

Data centric, information-driven; significant UArizona-AIL Influence;
AIL, HAL, DSAIL are all external funding-driven

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!

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 societal impact, NSF, media, or stakeholder (e.g., CyVerse, BI users) needs and interests.
  - Relate back to your bigger vision. Unlikely you can make big impact with one paper.

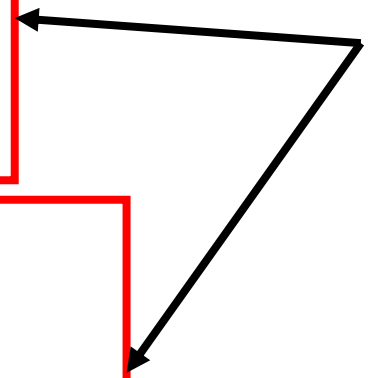
- If no one is researching a topic, ask why before pursuing it.
  - Lack of technology, not an interesting problem, not a real issue, etc.

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?
  - How is the method commonly evaluated?



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

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); keywords
  - **Business Journals:** See FT50 or UTD24 journal lists
  - **Application oriented CS (conferences):** WWW, ASONAM, KDD, ICDM, VLDB, AAAI, ACL

- Aggregate papers (use **PaperPile**, Google Scholar, **Connected Papers**) 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 – (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
    - **Columns include:** year (newer first), author, focus, data, method, tools, significant results
  - Key observations (answers to key questions regard data, coverage, approaches, etc.)
    - Red box to highlight key aspects of the table that lead to key observation and observations.
  - Transition to method or another related area of domain OR methodology – based on data characteristics, limitations with existing methods, etc.

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)
  - **CS Conferences: NeurIPS, ICLR,** ICDM, ICML, ASONAM, KDD, CIKM…
  - **CS Journals:** IEEE IS, IEEE TKDE, IEEE PAMI, IEEE TNNLS, ACM TOIS, IEEE TEM, ACM TIST…

- Aggregate papers (use PaperPile + 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

- **Structure in slides:**
  1. Justification for the method – based on data characteristics, tasks, or requirements
  2. Background of the method (where did it originate → class of methods)
  3. Key definitions and operations (math, diagram, key steps)
     1. Who has used the method before and for what application?
  4. Key observations (limitations in your context)
  5. Transition to another method or research gaps and questions

- Diagram of how method operates is very helpful!

M2: AI Lab Template – Research Gaps and Questions

- **Example domain specific research gaps (based on transitions between sections):**
  - 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 (that are important in your context/task)
  - 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.

- Research questions should be targeted and driven from the research gaps.
  - Common mistake – not carefully aligning the research questions with the research gaps.

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

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

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; see DL-ISR JMIS paper):**
  - **Application** → formulating 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.)

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.

- Do not use terms/approaches that have not been defined in the lit review!

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.

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

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.

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, HL
  - **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.

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)

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! See Open Source Tools Repository slide deck.

- 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)

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 (consistent with how it presented in earlier sections).
  - Discuss what outperformed. For what reasons? Compare with past methods and literature.
  - Illustrate 1-2 meaningful examples where your proposed method detected instances missed by best performing benchmarks

- **“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.

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)

M2: Introduction, Conclusion, Abstract, Title

- Introduce and motivate the topic (sell your work from the high-level):
  - 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 (specific task for the paper).

- 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 of the context (e.g., data samples from hacker forum) can go a long way!**
  - Not being clear what the contribution is → bullet points of the key novelties can make it very clear

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.

Outline of Bootcamp

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

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.

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.

M3: Common Errors for Slides

- **Common errors for slides:**
  - Inconsistent use of terms – if its in the diagram, make sure the following slides reflect the terms!
  - No slide numbers
  - Inconsistent use of font sizes and colors (e.g., 24 pt for one main bullet, but 16 pt for another main bullet)
  - No or non-descriptive captions for figures and tables
  - Not referring the tables and figures in text
  - Incorrect or inconsistent reference styles (in-text or in the references section)
  - …

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.

M3: Interacting with Faculty for Feedback

- **Common errors for papers:**
  - Not following the abstract length/structure requested by Dr. Chen
  - Inconsistent formatting of text/sections/spacing
  - Not tightly linking and describing each figure and table to the surrounding text
  - Transitions between major sections of lit review are missing
    - Between domain to domain review – why?
    - Between domain to method review – why? What data/method characteristics?
  - Research gaps and questions not consistent with the transitions between lit review.
  - Incorrect proportions of content (e.g., too much case study, not enough experiment)
  - Inconsistent use of terminology
  - Figure/table sizes/fonts inconsistently labeled or sized
  - …

M3: Interacting with Faculty for Feedback

- Slides and papers should be carefully proofread **before** providing to your faculty mentor (don't waste their time – they are very busy!).

- Common mistakes students make when interacting with faculty:
  - Expecting the faculty to tell the students "exactly what to do and how to do it"
  - Expecting the faculty to do it for you
  - Not responding to faculty emails professionally – respond quickly, sign your emails, and do not treat email like text messages.
  - Not keeping faculty up-to-date of progress, roadblocks, or if you are working with them.
  - Expecting the faculty to stay up late with you the night before presentations
  - Not giving enough lead time for faculty to review (need to provide one week, minimum)

M3: Writing Tips and Tricks

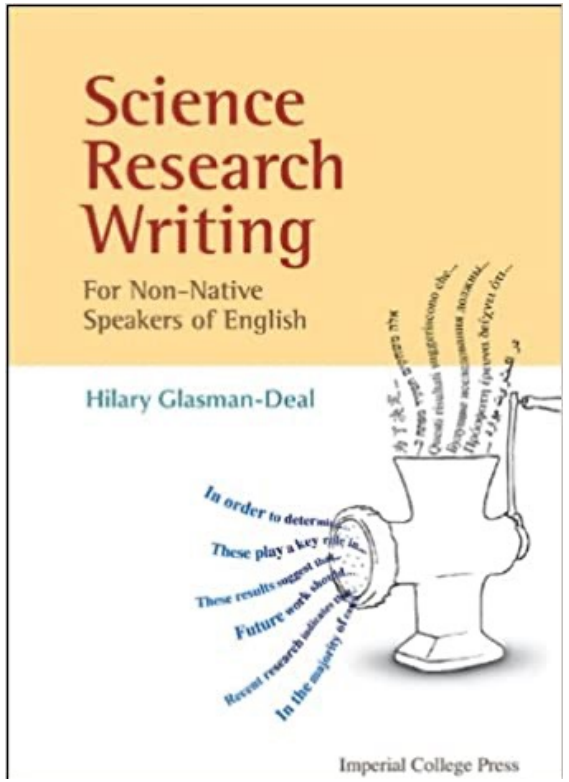
- Writing plays an important role in one's career → requires focus and effort.
  - Technical writing should be "precise and concise" and "short and sweet"

- **COGU:** Clarity, Organization, Grammar, Usage
  - Domestic speakers often struggle with **Clarity** (vague, non-specific terms, overuse of pronouns, colloquial) and **Organization** (inventing their own style, inconsistencies)
  - International speakers often struggle with **Grammar** (e.g., split infinitives, adjectival phrases, etc.) and **Usage** (overusing or misusing terms)

- **Good tools:**
  - **Writing quality:** Grammarly, ProWritingAid, The Writer's Diet
  - **Productivity:** RescueTime, pomodoro, planner

M3: Writing Tips and Tricks

- Omit needless words.
- Write everyday.
- Read your writing out loud.
- Do I really need this word/sentence 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.



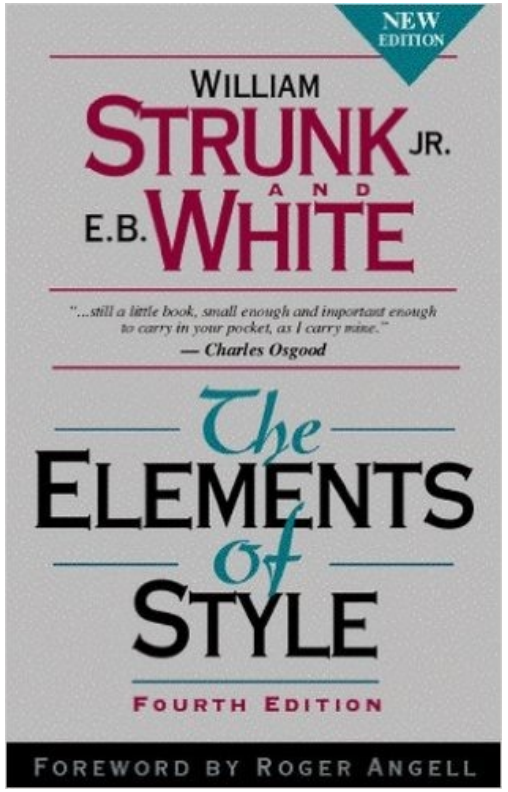


Figure Credits: Amazon

M3: Transitioning Slides to Quad Chart or a Poster

- Quad Charts are technical document used to briefly describe your work (examples in Bootcamp folder).
  - Four quadrants: Colorful, illustrative, highlights of your research focus, approach, results, impact; use at SFS job fair

- Sometimes you will be asked to present your work as a poster at conferences.

- 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

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.

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!

M3: Research Considerations – Managing Meetings

- Regular meetings with your advisor are critical.

- 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

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:** PaperPile, Google Scholar, Connected Papers
  - **Writing improvement:** Grammarly, Writefull
  - **Note Management and Collaboration:** Confluence, Notability, Evernote
  - **Public presence:** Google Scholar profile, DBLP, Semantic Scholar, personal website

- Keeping these up to date can help you quickly develop a suite of resources to rapidly advance your research, as well as help onboard new members quickly!

Outline of Bootcamp

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

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.).