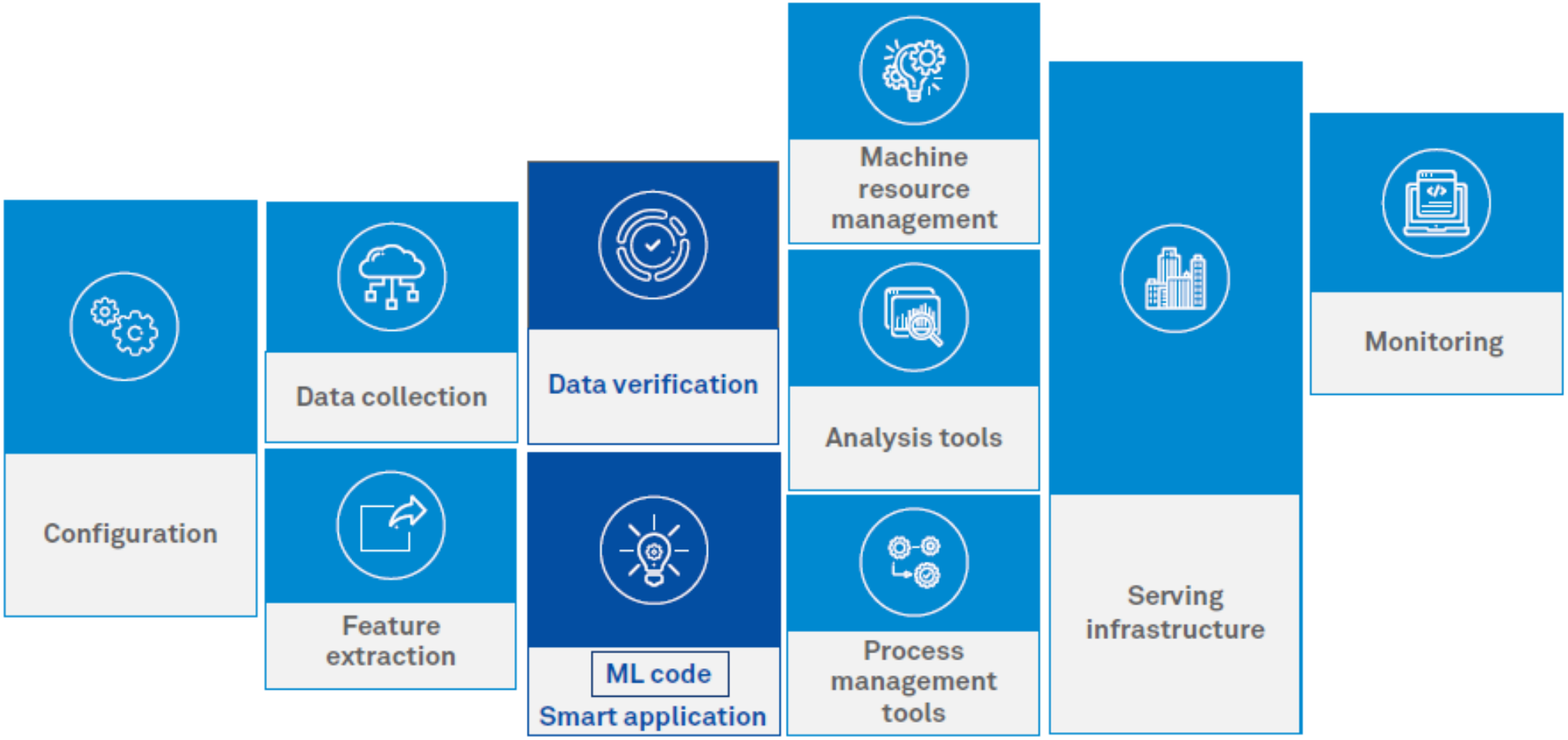
Testing and Application

Stephen Downes

December 13, 2021

Testing and Application



https://www.wipro.com/holmes/testing-of-ai-ml-based-systems/

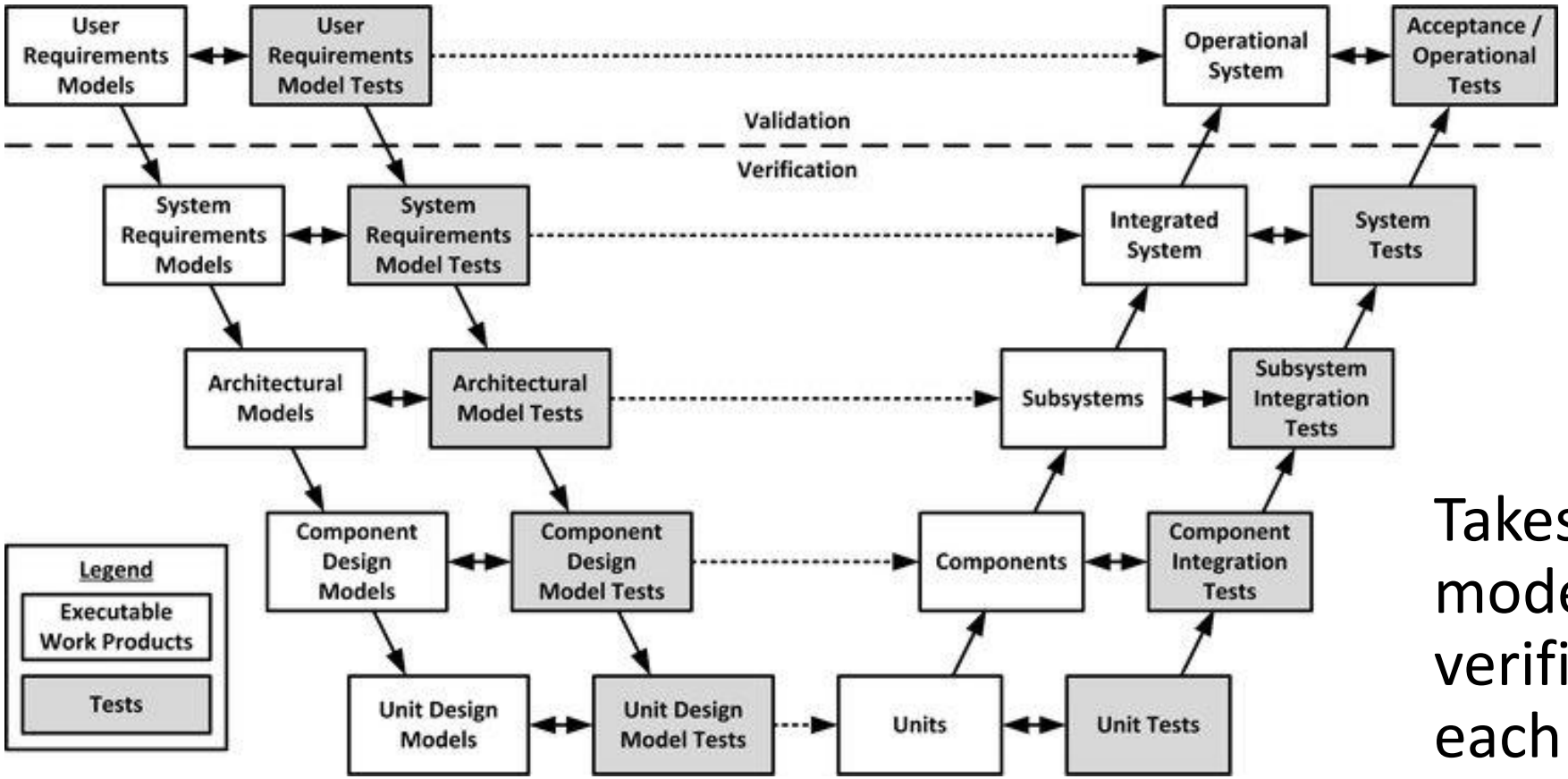
Testing: Objectives

- Define goals
- Identify KPIs
- Collect data
- Analyze data
- Perform test alternatives
- Implement changes

https://www.360logica.com/blog/web-analytics-testing-process-importance/

Image: Senuri Samindi https://senuri.medium.com/objectives-of-testing-4832858f6f7d

V Models for Testing

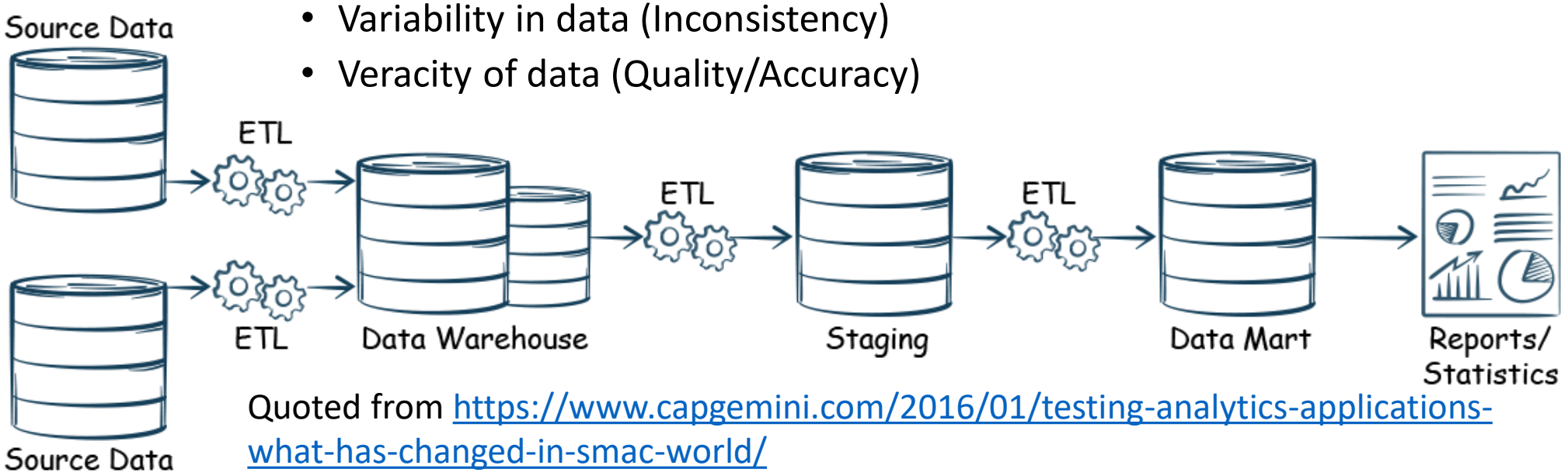


Takes the waterfall model and adds a verification step to each element

https://insights.sei.cmu.edu/blog/using-v-models-for-testing/

Data Testing

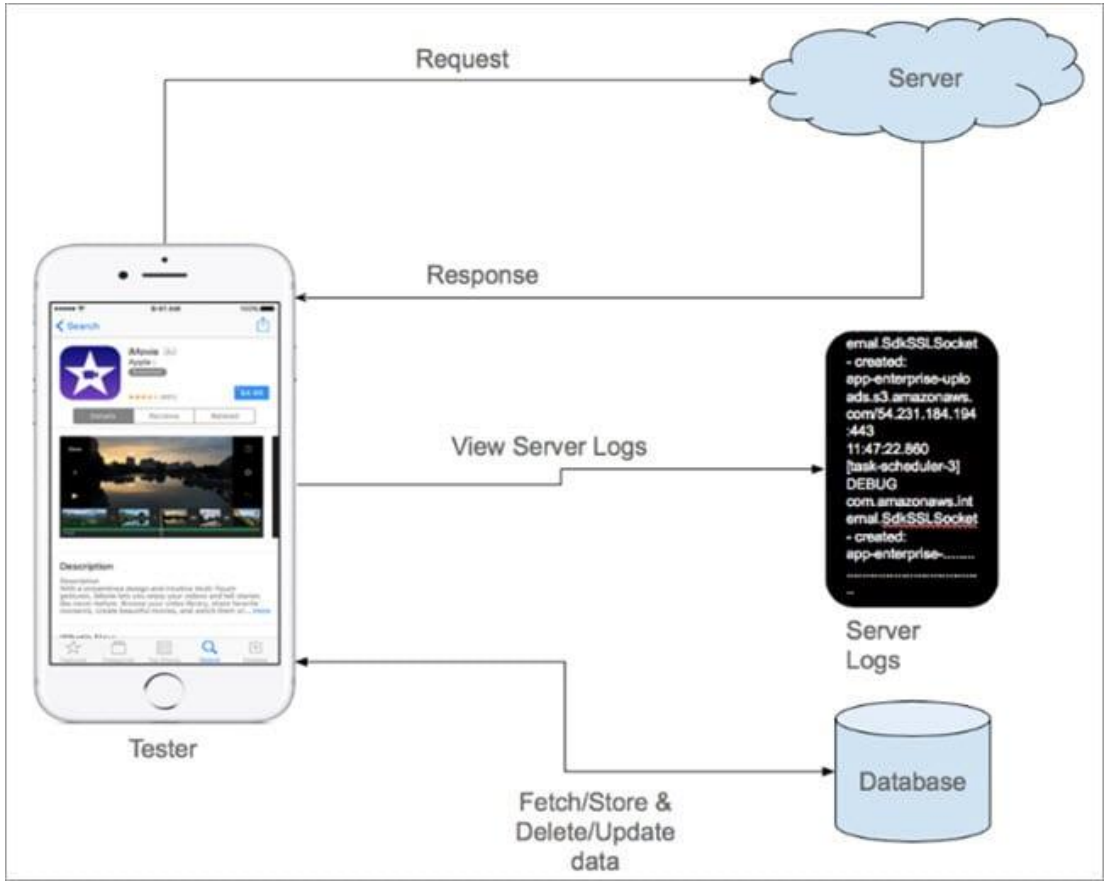
- The 6 Vs:
  - Data Volumes (Test for Semantics, Distributed processing, and scalability)
  - Data Variety (Test for visualization, schemas, and data federation)
  - Data Velocity (Test real-time, on the fly integration and on-demand storage)
  - Validity of data (apply rules and remove invalid data)
  - Variability in data (Inconsistency)
  - Veracity of data (Quality/Accuracy)



Quoted from https://www.capgemini.com/2016/01/testing-analytics-applications-what-has-changed-in-smac-world/

Request testing

- Ensuring that:
  - the correct data is being collected
  - the request to the AI is correct
  - The request is properly sent
- Validate:
  - Dynamic data, as in query parameters
  - For duplicate requests
  - Missing requests
  - Cross-browser functionality

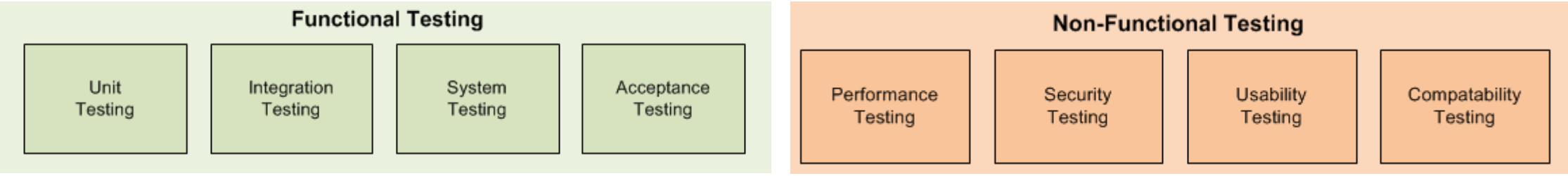


https://applitools.com/blog/guide-testing-automating-data-analytics-events-web-mobile/#h-different-ways-to-test-analytics-events Image: https://www.softwaretestinghelp.com/beginners-guide-to-mobile-application-testing/

Application Testing

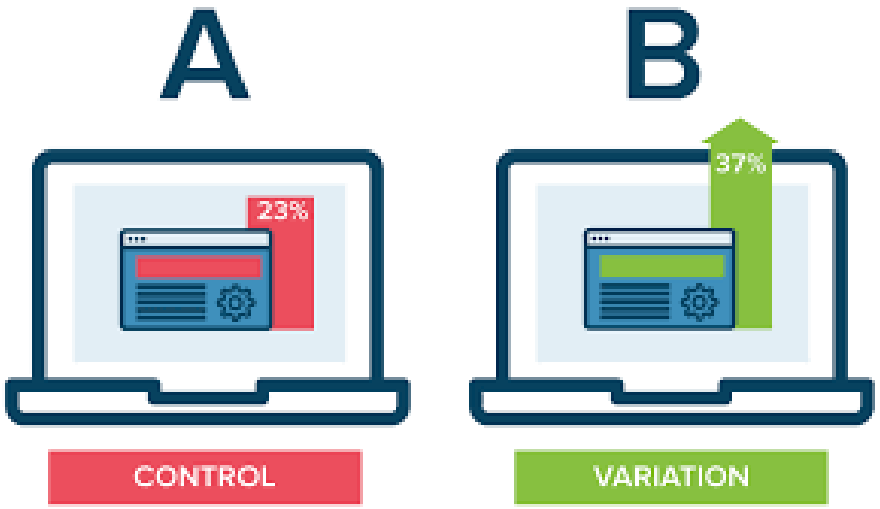
Analytics solutions should be tested for common testing techniques:

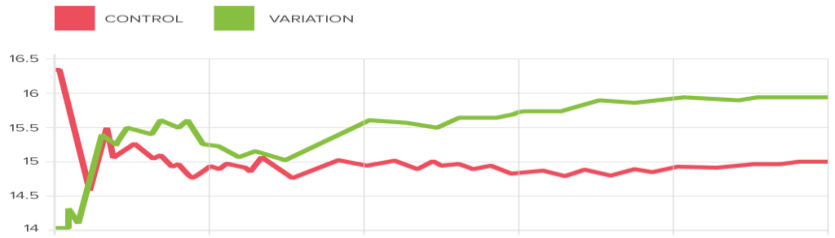
- Security Testing – authorization and authentication, availability of data
- Performance Testing – accuracy of data, and performance under high load
- Usability Testing – e.g. whether the application is providing right information
- Failover – to ensure data is available during critical failures



Quoted from https://www.capgemini.com/2016/01/testing-analytics-applications-what-has-changed-in-smac-world/ Image: https://www.inflectra.com/ideas/topic/testing-methodologies.aspx

Testing: A/B and Multivariate Testing





"This testing tells you about the performance of clickable elements of your web page. This testing approach allows gauging performance like rotating between two images or tweaking an important call to action and then acting on the results. This helps in getting a website with continuous performance improvement."

https://www.360logica.com/blog/web-analytics-testing-process-importance/
Image: https://www.optimizely.com/optimization-glossary/ab-testing/

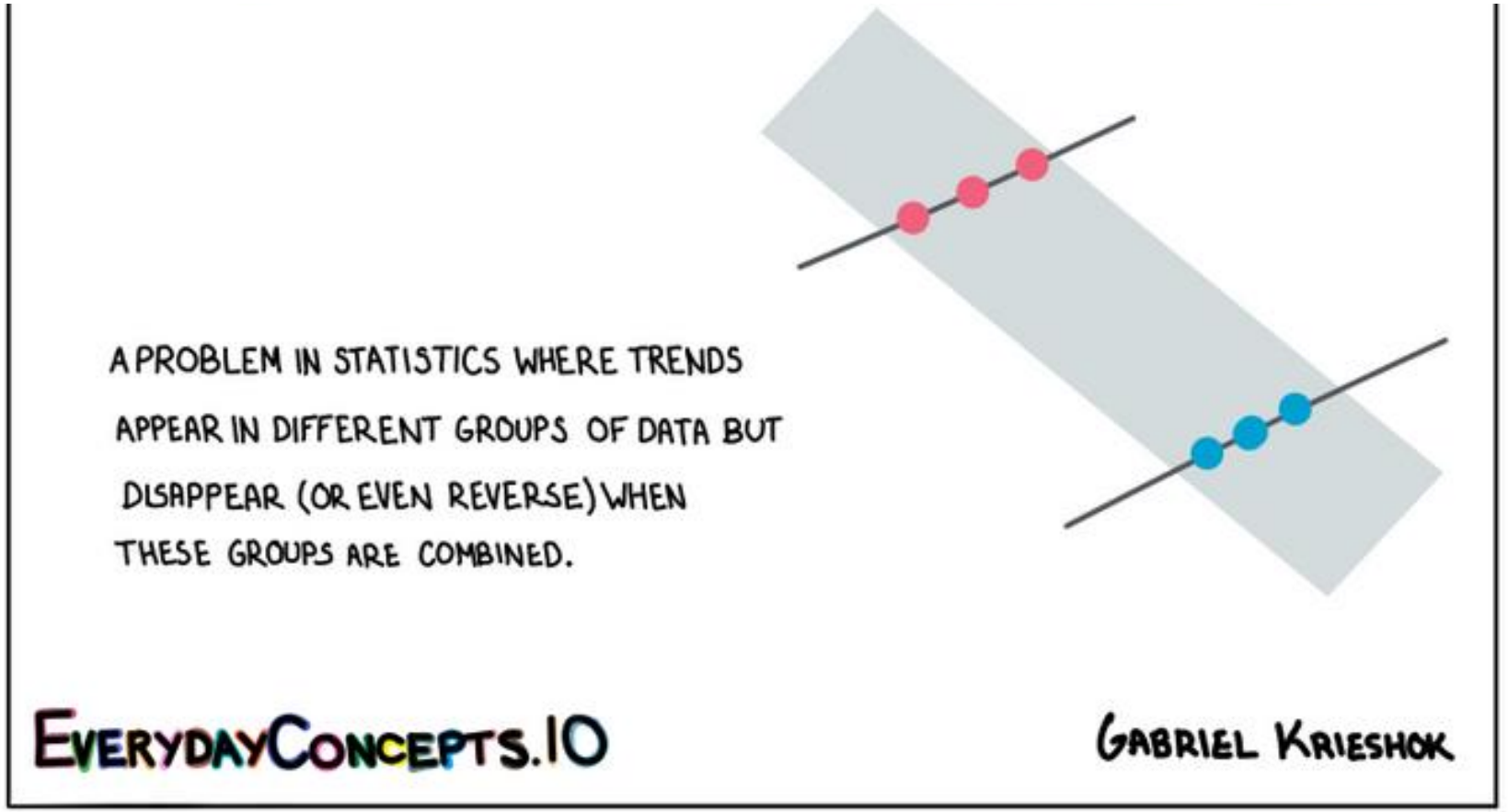
Testing: End Report

@BagmarAnand

By the time you're here, it's a bit too late

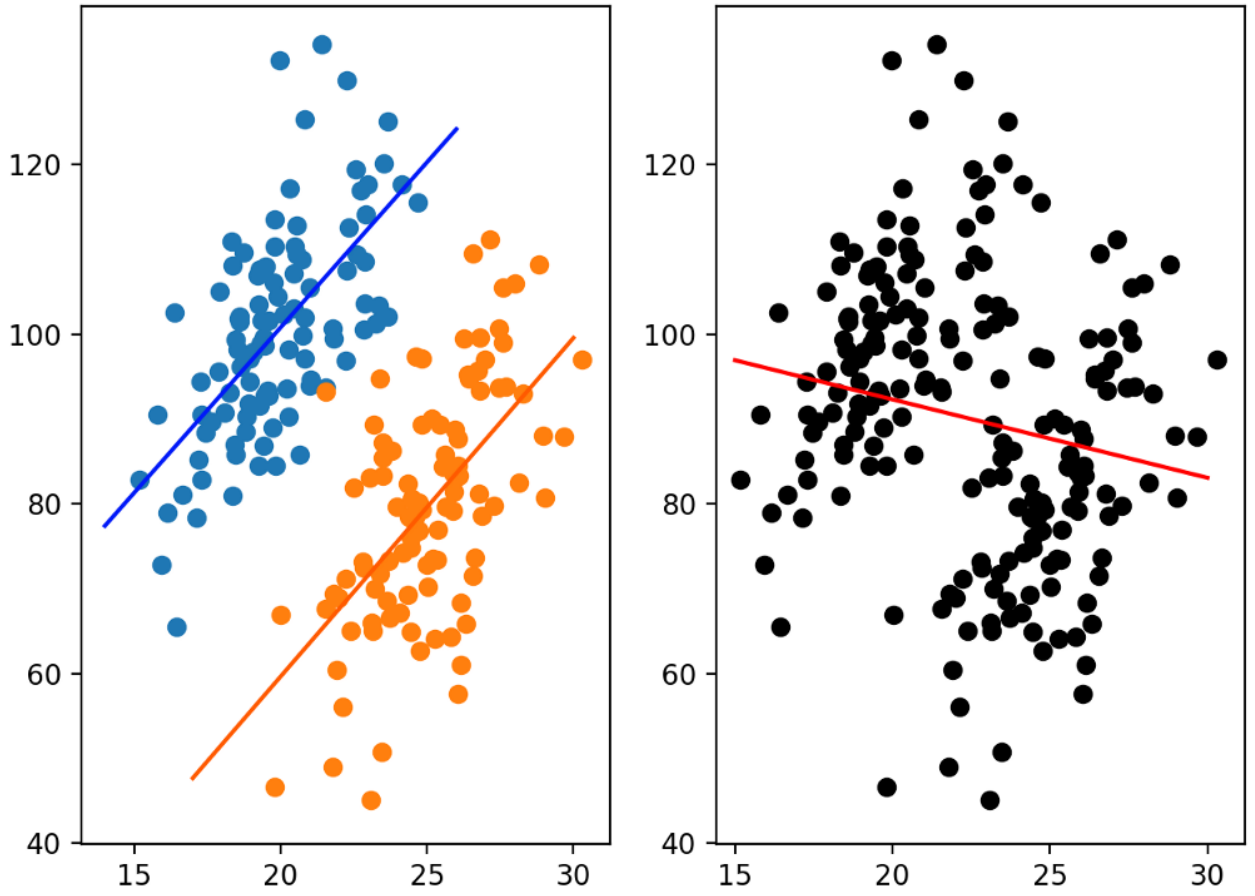
https://applitools.com/blog/guide-testing-automating-data-analytics-events-web-mobile/#h-different-ways-to-test-analytics-events

Simpson's Paradox



https://imgur.com/gallery/WQ6Pxy7

Scatterplots



https://imgur.com/gallery/WQ6Pxy7

Whose ‘F’ That Is

https://imgur.com/gallery/WQ6Pxy7

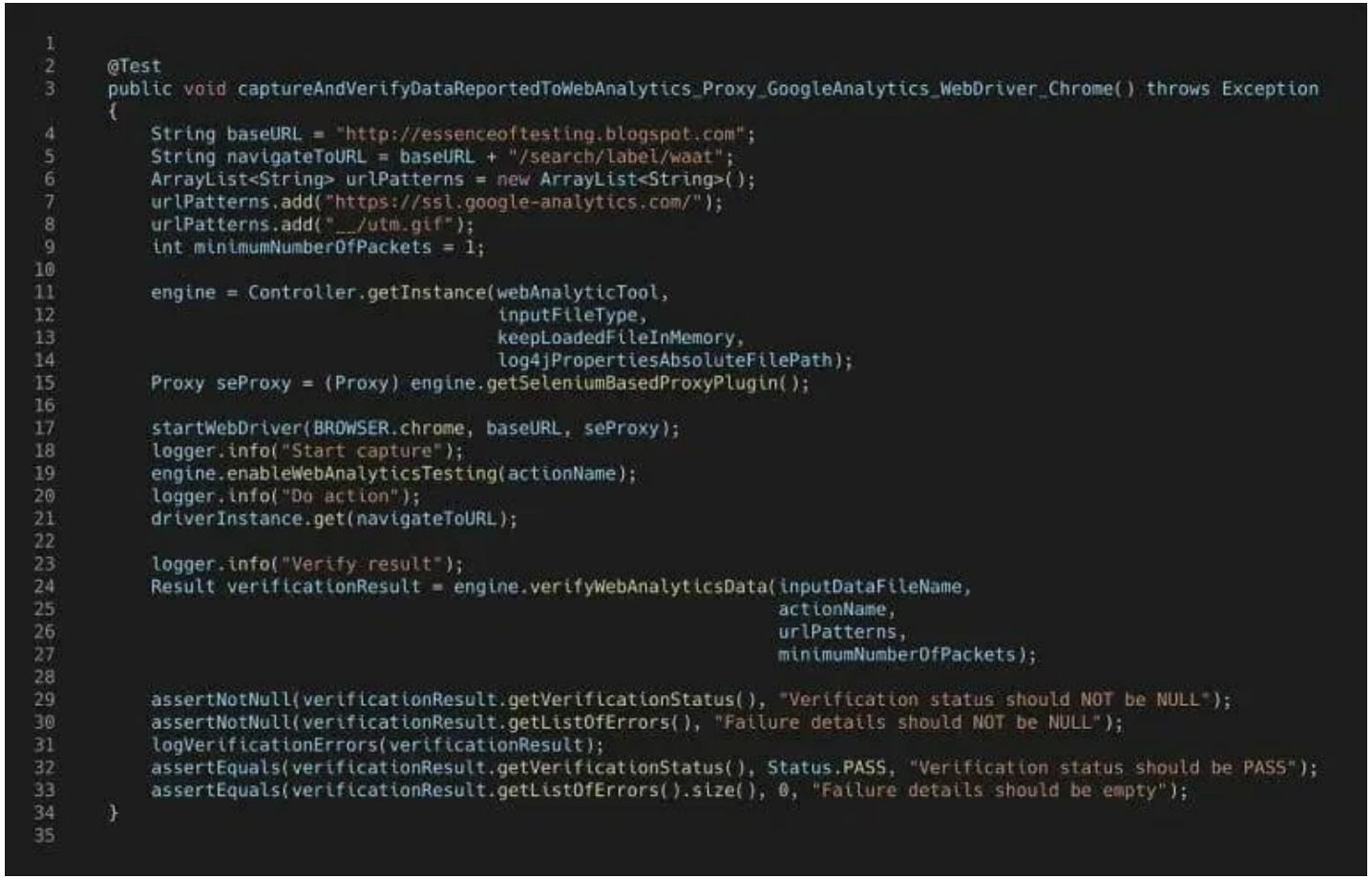
Testing: Automated Testing

Automated testing helps support agile development

https://applitools.com/blog/guide-testing-automating-data-analytics-events-web-mobile

https://medium.com/usa-today-network/automating-analytics-testing-25aa824c2e61

Video:
https://www.youtube.com/watch?v=al-_r8hVHmo



```
@Test
public void captureAndVerifyDataReportedToWebAnalytics_Proxy_GoogleAnalytics_WebDriver_Chrome() throws Exception
{
    String baseURL = "http://essenceoftesting.blogspot.com";
    String navigateToURL = baseURL + "/search/label/waat";
    ArrayList<String> urlPatterns = new ArrayList<String>();
    urlPatterns.add("https://ssl.google-analytics.com/");
    urlPatterns.add("__/utm.gif");
    int minimumNumberOfPackets = 1;

    engine = Controller.getInstance(webAnalyticTool,
                                    inputFileType,
                                    keepLoadedFileInMemory,
                                    log4jPropertiesAbsoluteFilePath);
    Proxy seProxy = (Proxy) engine.getSeleniumBasedProxyPlugin();

    startWebDriver(BROWSER.chrome, baseURL, seProxy);
    logger.info("Start capture");
    engine.enableWebAnalyticsTesting(actionName);
    logger.info("Do action");
    driverInstance.get(navigateToURL);

    logger.info("Verify result");
    Result verificationResult = engine.verifyWebAnalyticsData(inputDataFileName,
                                                              actionName,
                                                              urlPatterns,
                                                              minimumNumberOfPackets);

    assertNotNull(verificationResult.getVerificationStatus(), "Verification status should NOT be NULL");
    assertNotNull(verificationResult.getListOfErrors(), "Failure details should NOT be NULL");
    logVerificationErrors(verificationResult);
    assertEquals(verificationResult.getVerificationStatus(), Status.PASS, "Verification status should be PASS");
    assertEquals(verificationResult.getListOfErrors().size(), 0, "Failure details should be empty");
}
```

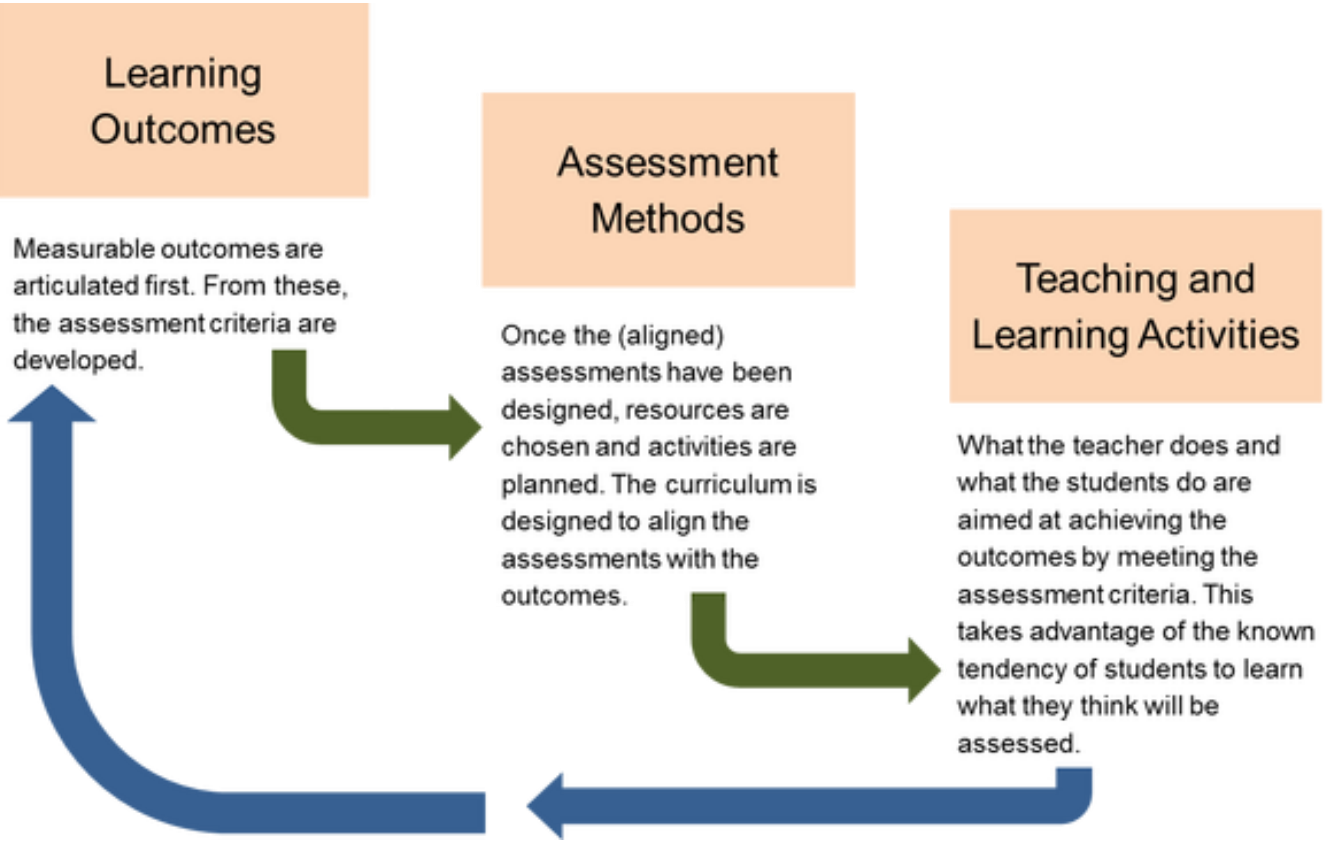
Standards For Validation And Transparency

- Rigor of the standards varies with the risks involved
- Data may be shared but identifiable data must be withheld
- Transparency key: “variables, such as age, sex, primary payer, inpatient utilization, and blood pressure, should be described.”
- Because of potential conflict of interest, “scientific peer review and independent validation are desirable”

https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048 https://iabtechlab.com/compliance-programs/

Outcomes Assessment

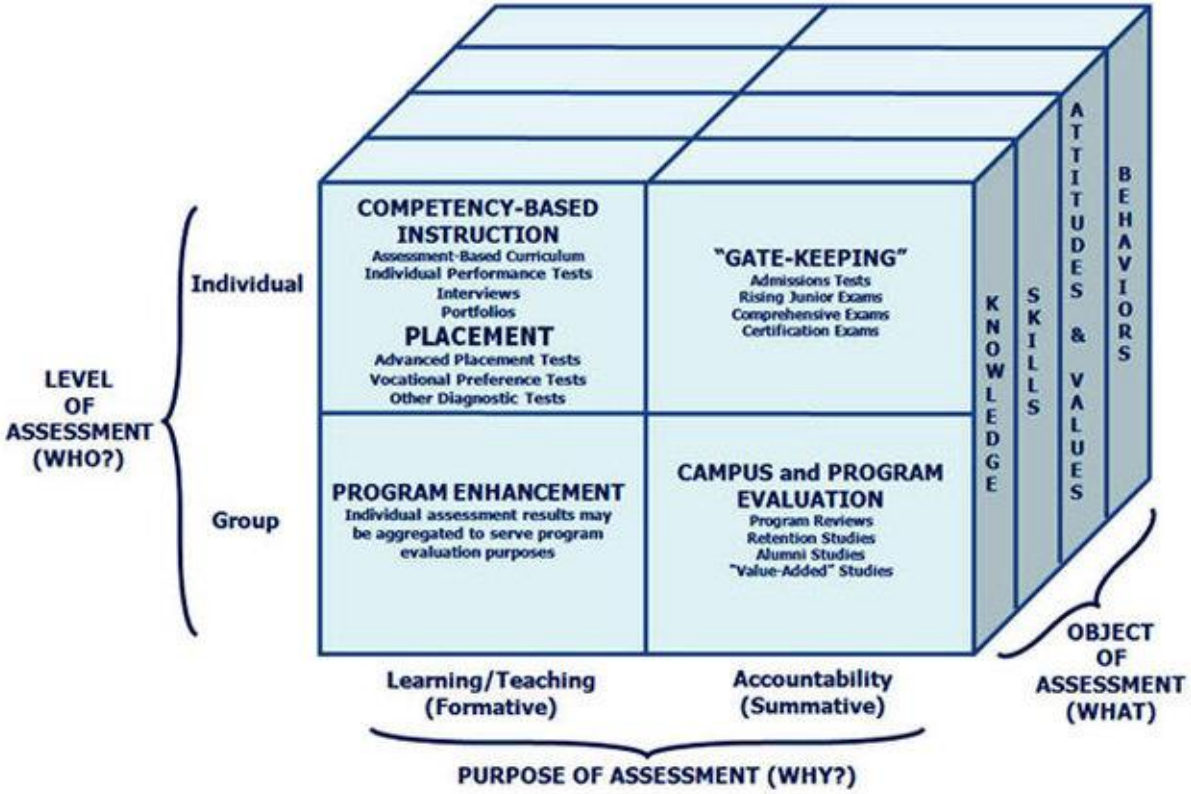
- Outcomes assessed should include:
  - ‘hard’ end points (readmissions, failures)
  - Secondary outcomes, such as care, trust, anxiety and activation
  - Provider-centered outcomes such as workflow and job satisfaction



https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048
Image: https://onedisruptiveeducator.com/2018/09/16/aligning-outcomes-assessments-and-activities/

Multiple Outcomes

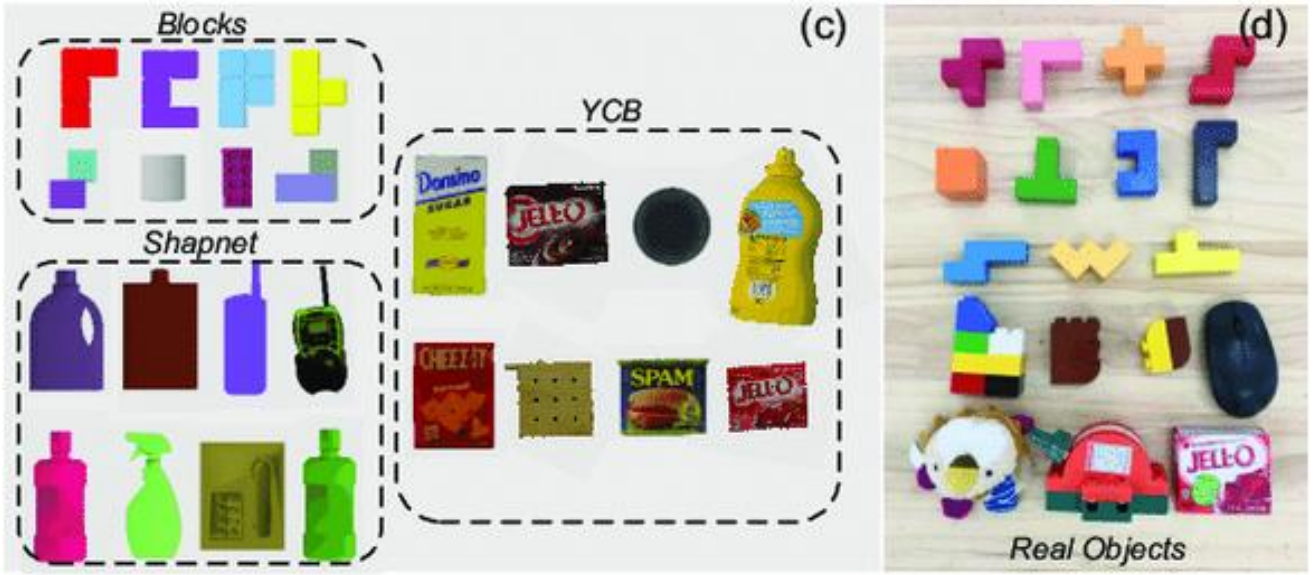
- Collect multiple outcomes
  - Each one may tell a different story
    - Can/should the model be localized, i.e. “monitored, refined, and reconfigured to local contexts”? (deBruijn, 2020)
  - Models may be valuable in one setting and process, but not others
  - Ensure student and instructor voices are heard



http://jeandowns.weebly.com/module-1.html

https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048

Conditions for Testing: Settings



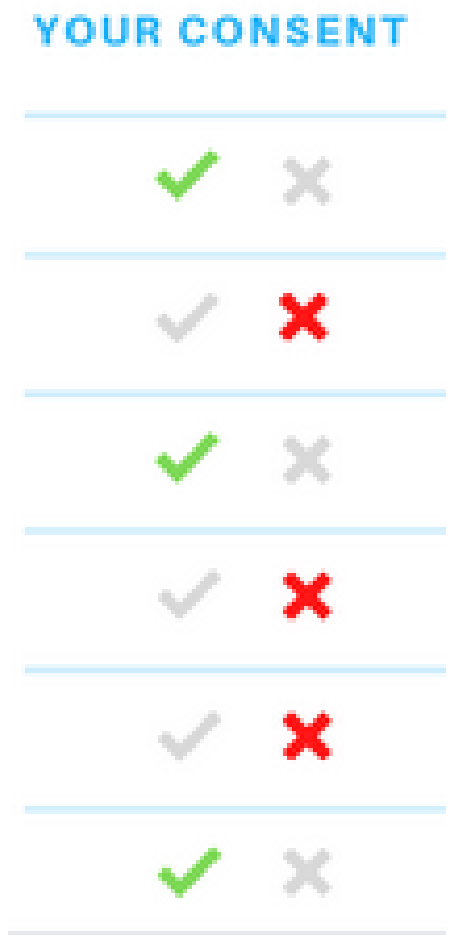
Testing The Model In Real-World Settings - the model would be tested under real-world conditions with adequate protections and precautions

https://www.researchgate.net/figure/Experimental-Setting-and-Testing-Objects-ab-Simulation-and-real-world-experimental_fig1_351289429

I. Glenn Cohen;Ruben Amarasingham;Anand Shah;Bin Xie;Bernard Lo, 2014. The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. Health Affairs Vol. 33, No. 7: Using Big Data To Transform Care. https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048

Conditions for Testing: Consent

- “It is unclear whether explicit consent to the use of personal data in predictive analytics is legally or ethically required.” Why?
    - patients are generally unaware if their physicians are using computerized decision aids
    - if people could opt out of existing allocation systems, their decisions might unfairly give them priority over other patients.
    - “the institutions under consideration should be required to explain whatever predictive analytics development and evaluation they are undergoing and the likely benefits and risks.”



https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048

Knowledge Mobilization

- Knowledge translation
  - coined by the Canadian Institutes of Health Research (CIHR) in 2000
  - "the exchange, synthesis and ethically-sound application of knowledge—within a complex system of interactions among researchers and users”
- Knowledge mobilization
  - “activities relating to the production and use of research results, including knowledge synthesis, dissemination, transfer, exchange, and co-creation or co-production by researchers and knowledge users”
  - “a term used to define the connection between academic research or creative works and organizations, people, and government”

https://ktdrr.org/ktlibrary/articles_pubs/ktmodels/
https://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/knowledge_mobilisation-mobilisation_des_connaissances-eng.aspx
https://www.uwinnipeg.ca/knowledge-mobilization/what-is-knowledge-mobilization.html

Application: Access

- Equitable Access
- There's a risk that not all benefit equally from the models
    - This is especially the case if there are licensing fees and other costs
    - Also, hat is the impact of fees, costs, patents and other factors restricting access to the model (even in some cases for people who provided some of the data) (deBruijn, 2020)
- "As a matter of fairness, those who contribute most to developing a model, including the patients who contribute their data, should proportionally enjoy its benefits"

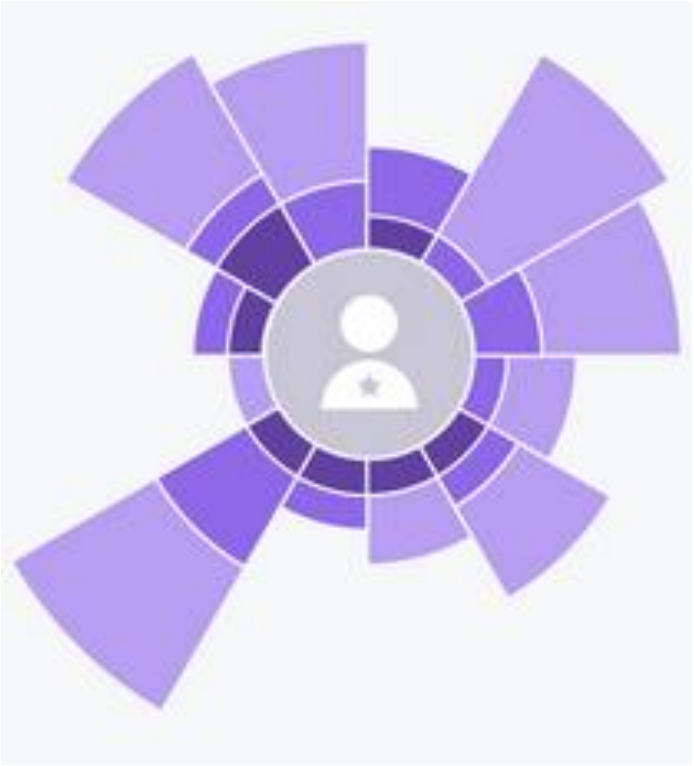
https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048
Also: https://medium.com/@AINowInstitute/ai-in-2018-a-year-in-review-8b161ead2b4e

Access and Power

- ‘When you have near omniscience, how you choose to apply that becomes a matter of importance’ - Stacy Higginbotham, TWIG, January 22, 2020. Eg. all the police resources will be applied to solve a robbery of a white well-off woman, but not the poor black woman.
- I would add - consider the case of missing and murdered indigenous women, where the police simply chose not to pursue the cases.

Application: The Human Element

James Clay (2020) writes, “We must not forget the human element of data and analytics. It’s not enough to deliver accurate analysis, predictions, and visualisations. Staff and students in universities and colleges need to be data literate to enable them to understand and act on that data. Appropriate and effective interventions will only be possible if staff and students are able to understand what is being presented to them and know what and how they could act as a result.”



Effective Learning Analytics: Using data and analytics to support students
Data does matter, and so does ethics
By James Clay January 24, 2020

Application: Implemetation

- Concern about flawed implementation
  - May be caused by zeal or pressure to cut costs
- May result from:
  - poorly constructed workflows
  - Insufficient consideration of client preferences
  - Inadequate checks and balances on machine decision-making
  - Used for ‘off-label’ and unproven uses

Mark Liberman observes, "End-to-end techniques, which eliminate human-defined layers like words, so that speech-to-text systems learn to map directly between sound waveforms and letter strings, are especially brittle." What this means is that AI is still (and for the foreseeable future) limited to specific context-insensitive domains.

Arun Narayanan et al., Recognizing Long-Form Speech Using Streaming End-To-End Models.
https://arxiv.org/abs/1910.11455 https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048

Application: Choice Architecture

Choice architecture - “To help the consumers of the model—both providers and patients—the model must present them with choices.”

Choice architecture is a concept of behavioral economics developed by Cass Sunstein and Richard Thaler, presented in their book [2]. The theory developed by them says that the decision-making process is impacted by how the options are shown.

https://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0048
https://towardsdatascience.com/machine-learning-and-the-architecture-of-choice-2cc69072a45b