



WWW.JBDTP.ORG

ISSN: 2692-7977

JBDTP Professional Vol. 1, No. 1, 2022

DOI: 10.54116/jbdtp.v1i1.16

---

## STRATEGIES FOR DEMOCRATIZATION OF SUPERCOMPUTING: AVAILABILITY, ACCESSIBILITY, AND USABILITY OF HIGH PERFORMANCE COMPUTING FOR EDUCATION AND PRACTICE OF BIG DATA ANALYTICS

---

**Jim Samuel**

Rutgers University

[jim.samuel@rutgers.edu](mailto:jim.samuel@rutgers.edu)

**Margaret Brennan-Tonetta**

Rutgers University

[mbrennan@njaes.rutgers.edu](mailto:mbrennan@njaes.rutgers.edu)

**Yana Samuel**

Northeastern University

[samuel.y@northeastern.edu](mailto:samuel.y@northeastern.edu)

**Pradeep Subedi**

University of Utah

[pradeep.subedi@utah.edu](mailto:pradeep.subedi@utah.edu)

**Jack Smith**

Marshall University

[smith1106@marshall.edu](mailto:smith1106@marshall.edu)

### ABSTRACT

There has been an increasing interest in and growing need for high performance computing (HPC), popularly known as supercomputing, in domains such as textual analytics, business domains analytics, forecasting, and natural language processing (NLP), in addition to the relatively mature supercomputing domains of quantum physics and biology. HPC has been widely used in computer science (CS) and other traditionally computation intensive disciplines but has remained largely siloed away from the vast array of social, behavioral, business, and economics disciplines. However, with ubiquitous big data, there is a compelling need to make HPC technologically and economically accessible, easy to use, and operationally democratized. Therefore, this research focuses on making two key contributions, the first is the articulation of strategies based on availability, accessibility, and usability (AAU) concepts for the demystification and democratization of HPC, based on an analytical review of Calibur, a notable supercomputer at its inception. The second contribution is a set of principles for HPC adoption based on an experiential narrative of HPC usage for textual analytics and NLP of social media data from a first-time user perspective. Both, the HPC usage process and the output of the early-stage analytics are summarized. This research study synthesizes expert input on HPC democratization strategies and chronicles the challenges and opportunities from a multidisciplinary perspective, of a case of rapid adoption of supercomputing for textual analytics and NLP. Deductive logic is used to identify strategies which can lead to efficacious engagement, adoption, production, and sustained usage for research, teaching, application, and innovation by researchers, faculty, professionals, and students across a broad range of disciplines.

**Keywords** *High performance computing, Education, Technology access, Supercomputing, Democratization, Big data, Artificial intelligence, Textual analytics, NLP*

## 1. Introduction

*Because HPC stands at the forefront of scientific discovery and commercial innovation, it is positioned at the frontier of competition—for nations and their enterprises alike...*  
(Ezell and Atkinson 2016)

Big data and artificial intelligences (AIs) are having a significant impact on business, work, governance, social interaction, and education. While big data and AIs hold tremendous potential for value creation and transformation of human life, their potential can only be realized through appropriate technological implementations. Specifically, significantly more powerful and scaled up data processing, networking, and data storage capabilities are required to harness the vast promise of AIs and big data. This necessitates the usage of technologies, which are popularly termed as “supercomputing,” and also as “high performance computing” (HPC), referring to the use of supercomputers or high performance computers for computing complex, voluminous, or iteration intensive calculations and analytics. The term “supercomputing” (or supercomputers) has been treated as being synonymous with HPC (or high performance computers), and has been parsimoniously described as a computer or a cluster of computers with far greater computing-memory-storage capabilities than a general computer, and as being “characterized by large amounts of memory and processing power” (George 2020). So also, HPC has been varying defined as being a “combination of processing capability and storage capacity” that can efficiently create solutions for “difficult computational problems across a diverse range of scientific, engineering, and business fields” (Ezell and Atkinson 2016), and also as being “massively parallel processing (MPP) computers” (Bergman *et al.* 2019). HPC can be classified as being homogeneous or heterogeneous, based on the use of similar or dissimilar processors (or memory, or similar HPC components) respectively, in its array of processors, such as homogeneous HPC with CPU arrays, and heterogeneous HPC with CPU and GPU arrays (Gao and Zhang 2016). Heterogeneous HPC can be used to improve effectiveness, speed, and also to gain additional energy savings. High performance computers can therefore be viewed as organized systems of high-powered, parallel structured computational capabilities, including extreme and diverse processing capabilities, general or task varied and scalable memory, scalable storage, grid or network or cloud based, and appropriate capabilities management interfaces with the potential to help solve vast and complex problems.

### 1.1 The Critical Need: Multidisciplinary HPC Applications

The compelling need for fostering HPC and HPC education has been well recognized by industry, government, and academia. However, most of these efforts have been largely isolated streaks, albeit with some progression, to a restricted set of traditionally computational domains. Given the explosive growth in quantity, diversity, complexity, granularity, and acceleration of data generation, it has become impossible to meaningfully depend on desktops, servers or standalone computers to create competitive value, and an increasingly large number of disciplines have begun HPC evaluation and adoption processes, to ensure that they remain competitive in progressively data and computation intensive environments (Fiore *et al.* 2018). The already steep trend towards HPC engagement and adoption can be expected to become stronger with the advent of new technologies and the identification of new opportunities (Bergman *et al.* 2019). An investigation by the Council on Competitiveness discovered that the vast majority of United States corporations with HPC capabilities had significant concerns about being able to hire persons with “sufficient HPC training,” and that there are no easy solutions because “... there aren’t enough faculty, researchers, educators, and professionals with the HPC skills and knowledge to fulfill the demand for talented individuals” (Lathrop 2016). There has been a sustained call over the past decade for opening up access to HPC/SC resources: “In the past decade high performance computing has transformed the practice and the productivity of science. Now this analytical power must be opened up to industry, to improve decision making, spur innovation and boost competitiveness” (Moran and O’Dea 2013). HPC has been treated as a critical catalyst for “inter- and trans-discipline breakthroughs” impacting the development of science and innovation globally (Mosin 2017). As illustrations, extant research has called for machine learning applications on varying types of information across a wide range of domains, include behavioral finance, social media analytics, textual data visualization and pandemic sentiment analysis, all of which are best implemented at scale using HPC resources (Samuel 2017; Samuel *et al.* 2018; Conner *et al.* 2019, 2020; Rahman *et al.* 2020). HPC is now a global phenomenon, and competitive advantage in many domains is associated with multidisciplinary HPC capabilities.

### 1.2 The Future of HPC Impact: Pervasive and Ubiquitous

Big data-driven AI holds the keys to global value creation (Samuel 2021; Samuel *et al.* 2022). Given the rapid growth of AI and the associated need to process big data, HPC has become one of the critical drivers of success for

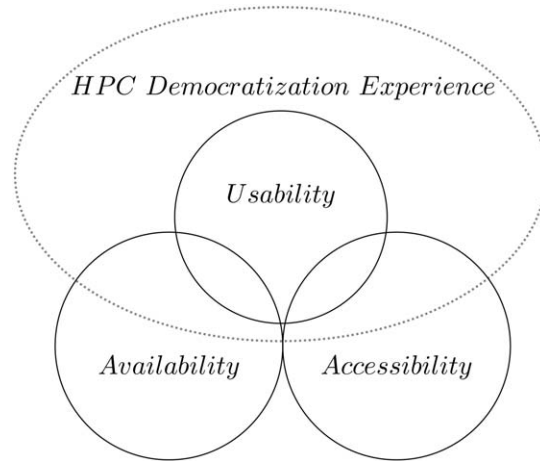


Figure 1: The HPC democratization concept.

research institutions, corporations, and nations. Ubiquitous AI and big data applications imply an equally expansive HPC requirement and influence. HPC ubiquity is a certainty, although end users will most likely not be required to interface with the technological complexity of HPC systems, just as electricity and internet are ubiquitous, without end users having to interface with systems managing electricity generation equipment or the movement of data packages and transmission protocols for the internet. Institutions, corporations, and countries which enable their constituents and stakeholders with easily usable HPC capabilities will possess a significant competitive advantage. HPC-driven competitive advantages will impact individuals, businesses, society, and nations, bearing the potential for significant socioeconomic impact. It is therefore of paramount importance to look closely at strategies for catalyzing future HPC thought leadership and competitiveness.

### 1.3 The Need: Democratized Supercomputing!

Democratized supercomputing, in the context of this study, refers to the opening up of HPC resources, freeing it from restrictive domain boundaries, and making it seamlessly available to all on an as-needed basis. Presently, even where HPC is available, it still remains inaccessible to many. Furthermore, even where access is provided, usability is restricted due to operational and skills barriers. Our quasi-phenomenological perspective based on the motivations driving the conceptualization, development, and deployment of Caliburn supercomputer at Rutgers, the State University of New Jersey, and an analysis of HPC usage for textual analytics and natural language processing (NLP) indicate that there are three key strategies that are necessary for democratizing HPC across domains: 1) availability strategy, HPC resources need to be built and distributed to ensure fair availability; 2) accessibility strategy, just the mere fact that HPC infrastructure exists in an institution or at a location does not ensure its accessibility, and therefore deliberate steps need to be taken to align and distribute available HPC resources in a manner so as to ensure HPC accessibility; and 3) usability strategy, availability and accessibility ensure that end users are empowered to access HPC resources; and yet these alone do not democratize or catalyze HPC utilization without the necessary dimension of ease-of-use. An effective HPC democratization initiative must include the three strategies of HPC Availability, HPC Accessibility, and HPC Usability, to ensure that capabilities are developed to achieve a maximized spectrum of benefits from HPC (Figure 1).

The rest of this paper is organized as follows. First, the study clarifies the multidisciplinary context, provides theoretical lenses from Information Systems (IS), and anchors HPC democratization discussion to the theories of technology adoption and usage. This is followed by an analytical and reflective narrative of the motivations and process for the acquisition and deployment of Caliburn, a supercomputer at Rutgers University. Availability, Accessibility, and Usability (AAU) strategies are then elaborated upon in subsequent sections. This is followed by a case analysis of HPC usage for NLP and key principles for sustained usage. The paper concludes with notes on implications, limitations, and a motivational conclusion.

## 2. HPC Engagement, Adoption, and Sustained Usage. Theoretical and Applied Considerations

### 2.1 The Future of HPC Relevance: Ubiquitous, Multidisciplinary, Transdisciplinary, and Interdisciplinary

Extant research meaningfully distinguishes between “multidisciplinarity,” “interdisciplinarity,” and “transdisciplinarity,” wherein “multidisciplinarity draws on knowledge from different disciplines but stays within their boundaries,” while interdisciplinarity “analyzes, synthesizes and harmonizes links between disciplines into a coordinated and coherent whole” and transdisciplinarity “integrates the natural, social and health sciences in a humanities context, and transcends their traditional boundaries” (Choi and Pak 2006; Alvargonzález 2011). We believe that such distinction is valuable. However, since this study does not delve into the nature of disciplinary research, but rather emphasizes the need for HPC to be used across disciplines, simultaneously drawing on knowledge and integrating lessons learned from the past, irrespective of discipline, hence we employ the word “multidisciplinary” in its broadest sense, inclusive of the properties of interdisciplinarity and transdisciplinarity.

### 2.2 HPC Engagement: Theoretical Basis

As with all technologies, there are critical drivers for HPC engagement, adoption and sustained usage. IS research studies have provided extensive insights into user engagement, adoption and sustained usage of technologies. The seminal, and in many senses foundation setting, technology acceptance model (TAM) theory validated and popularized the concepts of perceived usefulness and ease of use of technologies (Davis 1989). Subsequent studies updated TAM, including a broader theoretical basis with a unified perspective leading to the “unified theory of acceptance and use of technology” (UTAUT), which “highlights the importance of contextual analysis in developing strategies for technology implementation” (Venkatesh *et al.* 2003). Extant research has also demonstrated that the challenges of technology adoption and usage are subject to information facets, information complexity, equivocality of information, and information overload (Samuel 2016; Samuel and Pelaez 2017). Some technologies, such as blockchains, are pertinent to specific user categories, where generally dominant drivers of technology acceptance may have less relevance than factors such as “security, privacy, transparency, trust and traceability aspects” (Grover *et al.* 2019).

### 2.3 HPC Usage: From Theory to Practice

The generic usability and acceptance model (GUAM), in contrast to UTAUT, provides a significantly better explanation of “behavioral intention (72%) and technology use (63%)” for learning innovations (Obienu and Amadin 2021). This demonstrated that domain or discipline sensitive models have the potential to outperform generic adoption models like TAM or UTAT, due to variations on technological features and characteristics of user groups. Factors such as gender and social characteristics have also been shown to influence user engagement with technologies, such as the indication by prior research that women tend to weigh ease of use more strongly than men who tend to focus on perceived usefulness of the technology (Venkatesh and Morris 2000). Supportive technologies, such as anthropomorphic chatbots with human-like NLP and communication capabilities, can have a meaningfully positive impact on user perception of the usefulness of technology, and such supportive mechanisms can also support perceived ease of use (Rietz *et al.* 2019). Additional theories need to be evaluated to maximize the theoretical basis for HPC and supercomputing engagement. For example, flow theory which refers to the “the holistic sensation that people feel when they act with total involvement” and the state of “flow” where people experience becoming “absorbed in their activity,” akin to Chess players and gamers whose intelligences are fully engaged and focused (Csikszentmihalyi and Csikszentmihalyi 1992; Koufaris 2002; Csikszentmihalyi 2014). On the applied side, numerous AI research initiatives highlight the need for HPC in advancing research in multiple areas, such as natural language generation in the context of social media analytics (Garvey *et al.* 2021; Samuel *et al.* 2021). The present study presents an analysis of a first time users’ experience with HPC for textual analytics and NLP, to identify some useful principles that will help potential HPC users to move from theory and strategy to applications and practice.

## 3. Caliburn, a Story of Strategic Value Creation through ACI AAU

The story of Caliburn, the first supercomputer at Rutgers University and in the state of New Jersey, begins in 2011 with the creation of the Rutgers Discovery Informatics Institute (RDI<sup>2</sup>) by Dr. Manish Parashar, Distinguished Professor, Computer Science. His motivation was to establish a comprehensive and internationally competitive multidisciplinary Computational and Data-enabled Science and Engineering (CDS&E) institute at Rutgers University

that could catalyze and nurture the integration of research and education with advanced computing infrastructure (ACI). Parashar structured RDI<sup>2</sup> to provide ACI resources that were available, accessible and usable by offering technologies and expertise to academic researchers and companies that want to take advantage of ACI resources, but do not have the financial resources or expertise necessary to acquire these human and hardware resources. It was his vision to have a national level ACI at Rutgers University available to all.

### **3.1 Sensing the Need**

The importance and immediacy of having such ACI competency at Rutgers was further accentuated by the growing role of computation and data in all areas of science, engineering, and business, as well as current and future trends in ACI. These included disruptive hardware trends, ever-increasing data volumes, complex application structures and behaviors, and new first-order concerns such as fault-tolerance and energy efficiency. These trends are a result of the continued quest towards extreme scales in computing and data that is necessary to drive innovations in science, engineering, and other data intensive fields.

### **3.2 Innovation**

The CI developed by RDI<sup>2</sup> is innovative and provides researchers with global linkages to the national and international CI (e.g., XSEDE, OSG, OOI, LHC, iPlant, PRACE, EGI, etc.) that connects Rutgers with observational instruments, data streams, experimental tools, simulation systems, and individuals distributed across the globe. Overall, the impact of RDI<sup>2</sup> was a revolutionary advance in the scale and effectiveness of science and engineering research conducted at Rutgers and by academia and industry throughout the state.

### **3.3 Strategic Vision**

As a next step, it was critical that Rutgers develop and implement a bold strategic vision for an ACI ecosystem that was competitive at the national and international levels (Berman *et al.* 2013). This ecosystem had to provide researchers with cutting-edge computing and data handling capabilities, and students with necessary ACI exposure and training. In 2013, RDI<sup>2</sup> initiated a university-wide ACI strategic planning process with input from faculty across many disciplines at Rutgers. This resulted in a comprehensive plan, “Accelerating Innovation Through Advanced Cyberinfrastructure: A Strategic Vision for Research Cyberinfrastructure at Rutgers” (Berman *et al.* 2014). The plan called for strategic investment in ACI to drive innovation, improve research capabilities and productivity, and enhance faculty competitiveness. Two specific findings of the Rutgers ACI strategic plan were the need to deploy a nationally competitive advanced cyberinfrastructure and to establish a central Office of Advanced Research Computing at the University.

### **3.4 Availability**

Deploying a nationally competitive ACI required infrastructure investments in computing, mass storage, and high-speed/bandwidth digital communication that could provide state-of-the-art capacities and capabilities for Rutgers researchers and offer a competitive advantage among Big Ten peer institutions and beyond. However, the level of investment required to achieve this goal was significant. Fortunately, in 2013, the State of New Jersey announced the Higher Education Equipment Leasing Fund to support investments in cutting-edge equipment at the state’s higher education institutions. RDI<sup>2</sup> submitted a proposal entitled “Rutgers University Advanced Compute & Data Cloud” to establish a statewide ACI resource at Rutgers that could have far-reaching benefits for higher education institutions, industry, and state government. The state recognized the tremendous impact that this capability could have for its innovation economy, and the proposal was approved. RDI<sup>2</sup> was awarded \$10 million to purchase cutting edge ACI systems, the largest award given through this program.

### **3.5 Accessibility**

After two years of design and installation, Caliburn and companion system ELF were deployed in 2016 at Rutgers. The rationale for the two systems was the recognition of the potential limitations for many researchers who were inexperienced in a complex ACI such as Caliburn, or do not need its high-level capabilities. Thus, ELF, which had significantly more computing capability than what was currently available but not as complex or powerful as Caliburn, could be used by researchers as a first step in building experience and understanding of more sophisticated ACI. These systems provide a balanced advanced computational and data environment that contains a large-scale high-end compute engine, as well as significant co-located storage with embedded analytics capabilities. The Caliburn system was designed by SuperMicro in collaboration with the RDI<sup>2</sup> technical team. The latest in energy

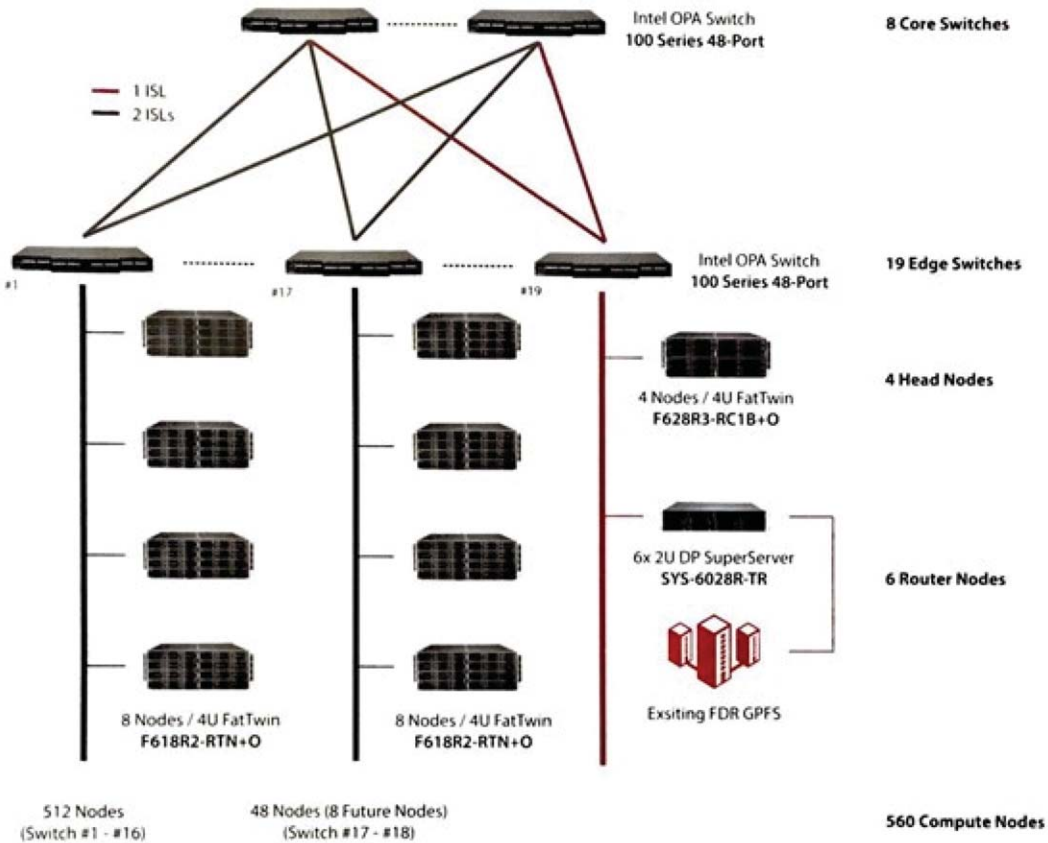


Figure 2: Overview of Caliburn architecture.

Table 1: Summary of Caliburn technical specifications

Caliburn	
Compute nodes	704
Cores	23,616
Memory	176 TB
Flash storage	200+ TB

efficient and space saving technologies were utilized, resulting in Caliburn using 90% less energy than standard indoor cooling units and is 80% space efficient. See Figure 2 for a diagram of the Caliburn architecture.

The Caliburn and ELF platforms are seamlessly accessible as a cloud service, providing researchers, students, industry and government across the state, with on-demand and pervasive access to these capabilities for research and instruction. Caliburn/ELF are also connected with high-speed networking to key national and international research/educational facilities. The overall platform is unique, and the most powerful academic system in the state. When it was commissioned in Summer 2016, Caliburn was ranked on the Top 500 list of computer systems worldwide as #2 system among the United States Big 10 institutions and #8 among all United States academic institutions, #50 among academic institutions globally, and #166 among all computer systems worldwide (Parashar 2019). Table 1 provides a quick overview for Caliburn’s technical specifications.

### 3.6 Usability

To further have ACI as available, accessible, and usable as possible at the university, RDI<sup>2</sup> began planning the creation of the first central Rutgers Office of Advanced Research Computing (OARC) in 2015. In Spring 2016, the

university's first Associate Vice President for Advanced Research Computing was appointed to continue the growth and development of university-wide ACI. The functions of this office are to provide strategic leadership, coordinate investments in ACI and related expertise, and catalyze and nurture cyberinfrastructure-enabled multidisciplinary research, all aimed at fostering a community of excellence in computing and data, empowering research, learning, and societal engagement and providing a competitive advantage to the Rutgers community and throughout the state and region. In 2019, management of the Caliburn and ELF systems was moved to OARC. These and other ACI resources are broadly available through this office.

### 3.7 Caliburn: Extending AAU with a Futuristic Vision

While Caliburn has been used for a wide range of multidisciplinary projects, one of the noteworthy initiatives which embodies the availability-accessibility-usability paradigm is the "Caliburn Supercomputing Awards." This initiative provides a pathway for scholars and academics outside of Rutgers to apply for HPC allocations, and thus expands the reach and impact of Caliburn; this is an example of a useful mechanism for the democratization of HPC: scholars and academics from institutions without HPC capabilities are now empowered in their research and thought leadership, which would otherwise be lacking. The process which was established by RDI<sup>2</sup> and continues with OARC, starts with OARC inviting proposals for "the allocation of computing resources on Caliburn," providing "high performance computing capabilities to academic researchers across the state to accelerate research programs that use or develop highly scalable computing applications." OARC also invites applicants in a second "startup" category, and these startup proposals "are provided as means to have full access with a limited time usage allocation" and are encouraged to be structured such that "they can be converted into awarded allocations during the next call for proposal cycle." Applications are limited to academic institutions in New Jersey under this program; this is something that OARC can go beyond, subject to availability of resources after meeting state-level needs, to provide a measure of access to meet the HPC needs of individual residents of New Jersey, nonacademic institutions, and academic outside of the state as well. For example, it would be of great value to New Jersey residents and students, if public libraries in New Jersey were empowered to provide interactive HPC demonstrations and interaction opportunities locally. Another avenue would be to explore relationships with specific institutions outside of the state to foster democratization of HPC resources.

## 4. The Expanding HPC Landscape: Notable Initiatives and Case Analyses

There are numerous existing efforts which in some form address the AAU paradigm. However, the focus is mostly on facilitating availability, and accessibility to a lesser extent. Usability tends to be left for the end user to wrestle with, with the help of "user guides," often leading to a loss of time and effort. The following section provides a brief overview illustrating some of the prominent HPC efforts.

### 4.1 HPC across Institutions and Disciplines

Advanced Cyberinfrastructure Research & Education Facilitators (ACI-REF): The Clemson-led ACI-REF program (NSF #1341935) advanced research computing through a network of Cyberinfrastructure (CI) Facilitators. This team is now co-leading Campus Research Computing Consortium (CaRCC). CaRCC is a NSF Research Coordination Network (NSF #1620695) and a follow-on to the ACI-REF project that addresses the huge growth in demand for local research computing, by sharing, collaborating, and developing best practices for research-facing, system-facing, software-facing, and stakeholder-facing CI professionals. CaRCC does not do large-scale workforce development for CI professionals itself, but most CaRCC institutions have now participated in Neeman's Virtual Residency Program. The ACI-REF Virtual Residency Program (VRP) was initiated by Henry Neeman, University of Oklahoma. He was a collaborator on the original ACI-REF proposal, and started the ACI-REF Virtual Residency Program (VRP) with an NSF CC-IEE grant (NSF #1440783) to provide national-scale CI Facilitator training. All the original ACI-REF institutions have participated in the VRP (subsection reference: Neeman *et al.*, 2016, 2018).

### 4.2 Facilitation of CI-Driven Research

XSEDE Campus Champions (CCs): There are more than 700 CCs, and these numbers are growing, at over 300 United States institutions helping their local researchers use CI, especially large scale/advanced computing. Most CCs perform CI facilitation activities, and CCs usually peer-mentor each other. The Champion community has: (a) a very active mailing list, where CCs exchange ideas and help each other solve problems; (b) regular conference calls for learning what's happening both among CCs and in national CI; and (c) major participation at national conferences like PEARC (e.g., 23% of PEARC'20 attendees were CCs). Many CCs have also participated in the VRP.

The Society of Research Software Engineering (SRSE) has 29 participating universities, supports Research Software Engineers (RSEs), focusing on reproducibility, reusability, and accuracy. The goal is to foster career paths for academic RSEs and ensure that they are recognized and rewarded. The United States Research Software Engineer Association (US-RSE) is the United States counterpart for SRSE. It has over 700 members. The United States Research Software Sustainability Institute (URSSI) has been funded by NSF from 2017 to 2021. Their goal is to design an institute on research software and to build the RSE community, in order to (i) improve how individuals and teams function, and (ii) advance research software and the STEM research it supports. Other efforts include initiatives such as the Supercomputing in Plain English (SIPE) workshop, which is an annual workshop on supercomputing (HPC) at Oklahoma University, run by Henry Neeman, using plain English to introduce fundamental issues of supercomputing as they relate to Computational and Data-enabled Science & Engineering. Internet2 and EDU-CAUSE also have programs to help enable CI Facilitators (subsection reference: Neeman *et al.* 2016, 2018).

### 4.3 NLP Case: Multidisciplinary HPC Education and Productivity

In addition to the analysis and study of institutional level initiatives, it is important to factor in individual perspectives of HPC usage from a multidisciplinary perspective. This subsection summarizes the key conceptual factors and implications for HPC usage from the lens of a first time HPC user for textual analytics (TAn) and NLP on social media data. TAn and NLP have been widely used for social media data analytics and a broad range of natural language sense-making efforts, including research on COVID-19, stock market, and public perception (Kretinin *et al.* 2018; Samuel *et al.* 2020a, 2020b). The narrative is based on a Caliburn allocation award for a TAn and NLP project; this narrative does not focus on the findings of the core research, but rather on the process employed by a new HPC user, and the associated learning curve. The analysis highlights how the Caliburn award served as an excellent example of CI availability and facilitation of capability, but was lacking in sufficient usability support for a non-Computer Science (CS) user.

#### 4.3.1 Caliburn Usage Case: HPC for Making Sense of TAn and NLP

The goals and motivation for HPC engagement were to make sense of a large data file of social media data, consisting of over 7 million records, which needed to be cleaned, explored, summarized, and analyzed for general and public sentiment insights. The analysis was initiated using R and Python, and associated software packages and libraries. This task, which was initiated in 2019, was beyond the capabilities of a high-powered desktop with 64 GB of RAM and mandated the use of HPC resources. A Caliburn HPC allocation award made this analysis possible from 2020, and the second phase of the project continues into 2021. The sentiment analysis and custom advanced data visualization methods used for analysis necessitated the installation of new packages and libraries on the HPC system.

#### 4.3.2 Navigating Caliburn: Initiation

The HPC engagement process involved remote access of Caliburn with a two-factor authentication process. Rutgers OARC had a very clear step-wise process for this, and the initial access process was smooth and efficient, thus indicating that the availability of CI HPC resources were well supported by an efficient Accessibility strategy. The challenges occurred on the usability level, and although issues such as data transfer were self-resolved by the researcher's own efforts, issues with running necessary software remained. The major challenges faced were on key dimensions of usability: (a) interface, the command line drive interface led to a long learning curve, and this could be mitigated by the use of open-source solution such as OpenOnDemand; (b) software installation, while Caliburn had existing tools, it was a laborious and iterative process to figure out optimal ways to install all required R packages and Python libraries; and (c) exporting and saving the analysis, especially the data visualizations in the required format. A qualitative estimate indicates that about 75% of man-hours invested into the analysis were spent of resolving usability issues; this indicates a significant challenge for new users and non-CS users of HPC. Two kinds of HPC engagement were utilized: running live jobs (smaller subsets of data, relatively low computational requirement) and running batch jobs (the intended purpose of HPC, using larger datasets, with higher computational requirements). In running batch jobs, an additional issue arose: development of standalone script for running complete analytical processes; this requires a new mindset as compared to live-interactive analytics processes and is described in further detail in the Caliburn usage process subsection below.

#### 4.3.3 Caliburn Usage Process: Tactical Summary

In our case, we utilized two kinds of HPC engagement: running live jobs and running batch jobs. Running live jobs involved "asking" Caliburn for access to a HPC node, where once access was provided, it was like using a Linux



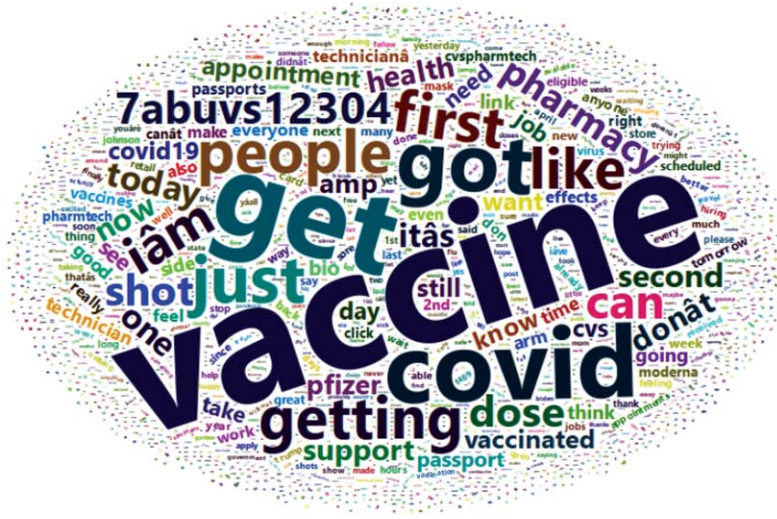


Figure 3: WordCloud on COVID-19 vaccine data.

machine via terminal. This use case is particularly suitable for analyses that require frequent user-intervention and employs smaller subsets of data with relatively low computational requirements. In contrast to live jobs, running batch jobs can be further divided into small scale jobs and large-scale jobs. Both batch cases require the development of a standalone script for running complete end-to-end analytical processes; this requires a new mindset as compared to live-interactive analytics processes. While the small-scale batch job is similar to running a live job from the perspective of storage and computational requirement, the benefit is that once the script is written, it does not require constant monitoring and interaction. Large-scale batch jobs are mainly designed for larger datasets with higher computational requirements, and it usually engages multiple nodes of the HPC machine. However, these large-scale batch jobs come with a steep learning curve on the efficient use of multiple nodes and cores of a HPC machine, and on end-to-end output-inclusive scripts that are required to run parallel analytics processes. Calibur was powerful enough to process a few million tweets fairly quickly. Large scale social media data analysis has become a powerful research paradigm, and has been used for a broad range of analyses, including public sentiment analytics (Ali *et al.* 2021). The processed data was then used for a variety of topical analyses on Calibur and on local machines, for running exploratory, predictive and machine learning models, and visualizations, including word-clouds on COVID-19 and vaccine data, as illustrated in Figure 3.

#### 4.3.4 Calibur Usage Case: Deductive Implications

The above described HPC usage analysis for TAN and NLP provided interesting principles and insights for HPC adoption. 1) Effective HPC Availability and Accessibility strategies led to the use of Calibur for business analytics purposes, employing TAN and NLP methods. This research would not have been possible with the Rutgers-OARC's Calibur award. 2) Usability strategies were limited to benefit expert HPC users and were not ready to support non-CS and new HPC users. 3) OARC support staff were well trained and provided critical assistance in addressing the barriers and issues with data transfers, software installation, and script required to enable TAN and NLP tools; this mitigated usability issues but still led to a significant increase in man-hours used for the analysis. 4) The absence of a GUI (graphic user interface) and user-friendly software installation process caused delays and discouraged additional creative research efforts, thus limiting the value created through for core research objectives. And 5) a successful HPC Usability strategy must implement appropriate user education and critical changes to the HPC process and system, to enable smooth and high impact multidisciplinary research.

#### 4.4 HPC Education and Productivity Principles

User education is critical for the success of any HPC Usability strategy. Education in HPC is an actively developing field with specialists from various disciplines participating in diverse initiatives. The development of specialized and purpose specific educational programs is therefore critical for meeting the ever-increasing demand for skilled HPC users (Connor *et al.* 2016). It has been shown that addressing the skill set gap is critical for meeting the needs in CI research workforce as well and interdisciplinary teaming helps foster the learning process in technological

domains (Choi and Kim 2017). Extant research has shown that interdisciplinary faculty are essential for successful implementation of HPC instruction (Neumann *et al.* 2017). Globally, emphasis on hands-on experiences and communications with international faculty is gaining significant prominence in HPC education (Sancho 2016). Several researchers call for a “holistic approach” to HPC training and education rather than focusing on a particular HPC ecosystem (Chaudhury *et al.* 2018). CI productivity can be maximized with an effective implementation of HPC Usability strategy.

## **5. HPC Democratization Strategies: AAU**

This leads us to summarizing a key contribution of this study: the articulation of AAU strategies for the demystification and democratization of HPC.

### **5.1 Availability**

This has been the first step in expanding HPC usage, and institutions across the nation and globally have been at the forefront of acquiring and developing supercomputing capabilities. An effective availability strategy consists of acquiring and developing CI, such that it caters to stakeholder needs for the current phase, while being scalable to accommodate larger workloads, and flexible to be developed for diverse workloads. In the age of cloud computing, needless to say, availability is not restricted by geography but bounded by network and access protocols. Once capacities were developed across many institutions, it was observed that underutilization was a problem due to accessibility issues. Multiple measures exist for the scale and scope of CI, which are essentially a summary of HPC technological components such as processors, memory, storage, and structure. Sufficient Availability is critical for HPC democratization. However, it is a costly error to believe that Availability will lead to maximized usage and optimal output.

### **5.2 Accessibility**

While some CI is built to cater to a very limited and specialized group of stakeholders, the Accessibility strategy is defined in reference to HPC capabilities at public organizations and academic institutions, where there is a need to cater to a broader need for HPC resources, as well as a responsibility to maximize the investment dollars. An effective Accessibility strategy consists of frameworks and processes that maximize the engagement of multidisciplinary stakeholders, plan for expanded user categories with prioritization of core stakeholders, such that underutilization of allocation is minimized. Accessibility strategies should be augmented with appropriate tracking and reporting of CI productivity and reach to evaluate the success of availability strategy. The minimization of underutilization of allocation of resources would serve as an indicator of success of an Accessibility strategy. There have been numerous creative efforts at expanding Accessibility and sharing of available HPC resources. Similar to Availability, expanded and equitable Accessibility is necessary for HPC democratization. However, Accessibility must lead to the productive, efficient, and effective state of Usability to ensure the highest likelihood of maximized usage and optimal output.

### **5.3 Usability**

There is a difference between optimal allocation of CI resources and optimal usage of CI resources. Maximizing the allocation of resources would be a measure for effectiveness of an Accessibility strategy, but that does not ensure optimal utilization and productivity at the end user level. Productivity maximization at the end user level not only requires good Availability and Accessibility strategies, but also a robust Usability strategy. Based on lessons learned from Caliburn, usage experiences, technology engagement theories, and a broader HPC landscape review, we describe a Usability strategy to maximize end-user level productivity. An effective HPC Usability strategy consists of a well-designed HPC system with multidisciplinary orientation, easy to use human interfaces, expert usage support, and general and discipline-specific applied HPC education. Productivity maximization at the end user level would serve as an indicator of success for Usability strategies. Basic measures, for example, could use averages of ratios of actual consumption to total allocation per user, across users in a category. Such ratios and measures would also serve as a check on over-allocation (excess availability at an individual end-user level) as well. An effective Usability strategy has been missing across many CI initiatives and addressing this strategy can lead to a remarkable increase in HPC productivity without additional expensive investments into increasing HPC availability. Based on our conversations with numerous CI experts and campus leaders, although there remains a need to improve Availability and Accessibility, yet the critical pathway to the highest likelihood of maximized usage and optimal output

is through the development and expansion of Usability factors; enhanced multidisciplinary Usability is the key for sustaining and ensuring the highest return on HPC investments.

## 6. Implications

HPC as an evolving technological, economical, and social multidisciplinary paradigm will have significant implication for human society, and this topic by itself will require a fair amount of research. This section does not attempt to discuss all potential implications, but is restricted to select issues most relevant to the current narrative, namely, the use of HPC productivity optimization strategies, HPC education as a key to HPC democratization and special issues and equal opportunity concerns with multidisciplinary HPC.

### 6.1 The Future of HPC Expansion: Decreasing Depth and Increasing Breadth

Moore's law predicted the doubling of transistors every two years, and the processor industry has experienced this curve till it reached the limits of physical properties; therefore, Moore's law will no longer be relevant to the future of new technologies, to the same extent that it has been in the past (Moore 1965, 1995; Schaller 1997; Theis and Wong 2017). HPC has entered into the early stages of a post-Moore era, and we have also seen significant advances in forms of massively parallel and hybrid forms of scalable computing. The underlying technologies have proven their worth and are well understood, leading to the greatest need: satisfy a broad range of domain-specific big data and voluminous algorithmic processing. We posit that the future is going to be relatively more strongly driven by an expansion the breadth of HPC applications across domains, rather than intensifying the vertical implementation of hardware improvements for marginal benefits in size and speed. This is supported by a growing demand for newer HPC applications, and relatively dwarfed demand for more sophisticated hardware. The success of cloud-based computing services such as Amazon Web Services (AWS) attest this perspective on current trends. Therefore, HPC availability and Usability strategies must be revised to cater to a broad range of disciplines, many of which will be traditionally non-computation-intensive disciplines. Development of HPC Usability strategies, and HPC education modules in particular, will have a significant impact on HPC democratization and productivity.

### 6.2 The Future of HPC Education: Modular, Applied, and Multidisciplinary

The goal here is to very briefly provide an impetus for HPC education, mostly focused on multidisciplinary curricular "modularization" and democratization. Notable initiatives include the "CyberAmbassadors" program, which is a 2017 CyberTraining project, focused on developing an open-source curriculum on interpersonal communication and mentoring skills for CI professionals. Similarly, SIGHPC Education Chapter is a virtual chapter of ACM's Special Interest Group in HPC and has merged with the IHPCTC (International HPC Training Consortia). They focus on developing best practices for HPC training, but don't provide such training themselves. The Linux Clusters Institute (LCI) holds workshops on HPC system administration, at introductory, intermediate, and advanced levels. These workshops have been extremely successful, typically attracting 20–40 HPC system administrators per workshop. LCI focuses on system-facing CI professionals, not researcher-facing. The Carpentries is an international volunteer organization that has run 2300+ hands-on workshops on research computing skills that so far have served 56,000+ researchers at 250+ institutions worldwide. They have demonstrated the effectiveness of "training the trainers" of researchers in informal education at large scale, with an emphasis on technical skills and pedagogy, and not on training CI Facilitators. The Coalition for Academic Scientific Computation (CASC): Members of this non-profit organization are primarily United States institutional CI leaders, plus some national CI leaders. CASC's focus closely aligns with CI leadership. CASC is not budgeted for or positioned to take on a major teaching or training role, but significant peer mentoring emerges from CASC activities. Science Gateways Community Institute (SGCI): the SGCI offers workforce development via internships, mentoring, and travel funding to conferences for graduate and undergraduate students, connections to the Young Professional Network, and support for gateway-related career paths (subsection reference: Neeman *et al.* 2016, 2018). In spite of many such initiatives, there is a critical need to evaluate Usability strategy-oriented education. A number of these initiatives cater to education pertaining to Availability and Accessibility and not to the proportionately higher need for multidisciplinary Usability training. This leaves a huge gap in educational needs being addressed for implementing effective HPC Usability strategies.

### 6.3 HPC Education: Special Issues and Equal Opportunity

Some of the challenges in the practice of HPC will be associated with addressing bias, such as gender bias in technology resulting in a low representation of women in HPC practice. Although there is very limited research done on the issue of gender and HPC, it is safe to assume that attracting and retaining women in HPC practice is going to be a challenge. This

is already so for the fields of STEM and especially CS, which is affected the most by a significant underrepresentation of women (Ehrlinger *et al.* 2018). Studies conducted on the topic of female underrepresentation found that less than 20% of the technological workforce is estimated to be women” (Frachtenberg and Kaner 2020). The adverse implications of having so few women in HPC are numerous and significant (Frantzana 2019). Extant research has emphasized the need to develop gender-specific learning strategies which accommodate women learners in technology disciplines (Samuel *et al.* 2020c). These issues will need to be addressed as aspects of the Usability strategy to ensure fair and balanced HPC democratization without bias, facilitating equal opportunities to persons in all categories.

## 7. Conclusion

This study identifies three key strategies for HPC democratization and important principles for catalyzing multidisciplinary HPC productivity. This research thus provides critical ideas and motivations for promoting HPC based research, applications and innovation in traditionally non-CS and non-computation-intensive disciplines and domains. We believe that this is a vital need for the current decade, and anticipate that this study will contribute to the body of knowledge that will influence HPC education policy in the future. We boldly propose and call for an increased emphasis on Accessibility and Usability strategies: every institution of higher education must ensure some measure of access to HPC resources through partnerships and networks, such as XSEDE, for their faculty and students. The responsibility for such efforts must be shared between those who “own” CI resources and those who need it.

“HPC is becoming a major driver for innovation offering possibilities that currently we cannot even evaluate or think about” (Puertas-Martín *et al.* 2020). We have emphasized that the dimension which needs the most attention is multidisciplinary HPC education within the HPC Usability strategy. We posit that undergraduate and graduate programs across disciplines must contain courses with HPC concepts and application modules, such as HPC lessons in IS courses for undergraduate business programs and in MIS courses for graduate business and relevant MS programs. Furthermore, workshops and interactive virtual education modules can be used for topic and discipline specific training. An appropriate HPC-Usability strategy and forward looking HPC education modules will ensure demystification of HPC and popularize its usage for innovation and value creation, by a broader range of students from multiple disciplines, and thus nurture the future HPC and AI workforce.

Furthermore, it is obvious that institutions and corporations with fewer resources, lacking research and technological infrastructure development funding, are at a disadvantage when it comes to HPC usage for research, teaching, and practice. An invigorated vision to democratize HPC reach into the smallest of institutions, going beyond boxed-in notions of traditionally bounded HPC domains, will maximize the return on investment for CI resources, as well as promote the noble and forward looking cause of better educating a robust future technological workforce.

**Acknowledgments:** We would like to thank Dr. Manish Parashar for his input and support for this research. We would also like to thank the OARC at Rutgers University for granting the two Caliburn allocation awards which facilitated this research and the RUCI lab (Rutgers Urban and Civic Informatics Lab; <https://rucilab.rutgers.edu/about-ruci/>).

## References

- Ali, G. G. Md. Nawaz, Md. Mokhesur Rahman, Md. Amjad Hossain, Md. Shahinoor Rahman, Kamal Chandra Paul, Jean-Claude Thill, and Jim Samuel. 2021. “Public Perceptions of COVID-19 Vaccines: Policy Implications from US Spatiotemporal Sentiment Analytics.” *Healthcare* **9**, no. 9: 1110. doi: [10.3390/healthcare9091110](https://doi.org/10.3390/healthcare9091110)
- Alvargonzález, David. 2011. “Multidisciplinarity, Interdisciplinarity, Transdisciplinarity, and the Sciences.” *International Studies in the Philosophy of Science* **25**, no. 4: 387–403. doi: [10.1080/02698595.2011.623366](https://doi.org/10.1080/02698595.2011.623366)
- Bergman, Keren, Tom Conte, Al Gara, Maya Gokhale, Mike Heroux, Peter Kogge, Bob Lucas, Satoshi Matsuoka, Vivek Sarkar, and Olivier Temam. 2019. “Future high performance computing capabilities: Summary report of the Advanced Scientific Computing Advisory Committee (ASCAC) subcommittee.” Technical Report. Washington, DC: United States Department of Energy Office of Science.
- Berman, Helen, Margaret Brennan, and Manish Parashar. 2013. “Accelerating Innovation through Advanced Cyberinfrastructure: The Urgent Need for Strategic Planning for Cyberinfrastructure at Rutgers.” In *Rutgers Discovery Informatics Institute Report*. Rutgers: The University of New Jersey.
- Berman, Helen, Manish Parashar, Don Smith, and Margaret Brennan. 2014. “Accelerating Innovation through Advanced Cyberinfrastructure: A Strategic Vision for Research Cyberinfrastructure at Rutgers.” In *Rutgers Discovery Informatics Institute Report*. Rutgers: The University of New Jersey.
- Chaudhury, Bhaskar, Akshar Varma, Yashwant Keswani, Yashodhan Bhatnagar, and Samarth Parikh. 2018. “Let’s HPC: A Web-Based Platform to Aid Parallel, Distributed and High Performance Computing Education.” *Journal of Parallel and Distributed Computing* **118**: 213–32. doi: [10.1016/j.jpdc.2018.03.001](https://doi.org/10.1016/j.jpdc.2018.03.001)

- Choi, Bernard C. K., and Anita W. P. Pak. 2006. "Multidisciplinarity, Interdisciplinarity and Transdisciplinarity in Health Research, Services, Education and Policy: 1. Definitions, Objectives, and Evidence of Effectiveness." *Clinical and Investigative Medicine* **29**, no. 6: 351–64.
- Choi, Ji-Eun, and Hale Kim. 2017. "Vertically Integrated Projects (VIP) at Inha University: The Effect of Convergence Project Education on Learning Satisfaction." *2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 436–43. IEEE.
- Conner, Cherilyn, Jim Samuel, Andrey Kretinin, Yana Samuel, and Lee Nadeau. 2019. "A Picture for the Words! Textual Visualization in Big Data Analytics." In *Northeast Business and Economics Association: NBEA Annual Proceedings-46*, Rhode Island, USA, 37–43.
- Conner, Cherilyn, Jim Samuel, Myles Garvey, Yana Samuel, and Andrey Kretinin. 2020. "Conceptual Frameworks for Big-Data Visualization: Discussion on Models, Methods and Artificial Intelligence for Graphical Representations of Data". In *Handbook of Research for Big Data: Concepts and Techniques*. USA: Apple Academic Press.
- Connor, Carolyn, Amanda Bonnie, Gary Grider, and Andree Jacobson. 2016. "Next Generation HPC Workforce Development: The Computer System, Cluster, and Networking Summer Institute." *2016 Workshop on Education for High-Performance Computing (EduHPC)*, 32–9. IEEE.
- Csikszentmihalyi, Mihaly. 2014. "Play and Intrinsic Rewards." In *Flow and the Foundations of Positive Psychology*, 135–53. San Diego: Springer.
- Csikszentmihalyi, Mihaly, and Isabella Selega Csikszentmihalyi. 1992. *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge: Cambridge University Press.
- Davis, Fred D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* **13**, no. 3: 319–40. doi: [10.2307/249008](https://doi.org/10.2307/249008)
- Ehrlinger, Joyce, E. Ashby Plant, Marissa K. Hartwig, Jordan J. Vossen, Corey J. Columb, and Lauren E. Brewer. 2018. "Do Gender Differences in Perceived Prototypical Computer Scientists and Engineers Contribute to Gender Gaps in Computer Science and Engineering?" *Sex Roles* **78**, no. 1–2: 40–51. doi: [10.1007/s11199-017-0763-x](https://doi.org/10.1007/s11199-017-0763-x)
- Ezell, Stephen J., and Robert D. Atkinson. 2016. "The Vital Importance of High-Performance Computing to us Competitiveness." *Information Technology and Innovation Foundation* **28**: 1–58.
- Fiore, Sandro, Mohamed Bakhouya, and Waleed W. Smari. 2018. "On the Road to Exascale: Advances in High Performance Computing and Simulations—An Overview and Editorial." *Future Generation Computer Systems* **82**: 450–458.
- Frachtenberg, Eitan, and Rhody Kaner. 2020. "Representation of women in high-performance computing conferences." Technical report, EasyChair.
- Frantzana, Athina. 2019. "Women's Representation and Experiences in the High Performance Computing Community." <https://era.ed.ac.uk/bitstream/handle/1842/36127/Frantzana2019.pdf?sequence=1>
- Gao, Yuxiang, and Peng Zhang. 2016. "A Survey of Homogeneous and Heterogeneous System Architectures in High Performance Computing." *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, 170–75. IEEE.
- Garvey, Myles D., Jim Samuel, and Alexander Pelaez. 2021. "Would You Please like my Tweet?! An Artificially Intelligent, Generative Probabilistic, and Econometric Based System Design for Popularity-Driven Tweet Content Generation." *Decision Support Systems* **144**: 113497. doi: [10.1016/j.dss.2021.113497](https://doi.org/10.1016/j.dss.2021.113497)
- George, Nathan. 2020. "Literature Review and Implementation Overview: High Performance Computing with Graphics Processing Units for Classroom and Research Use." Preprint, submitted May, 2020. *arXiv preprint arXiv:2005.07598*.
- Grover, Purva, Arpan Kumar Kar, Marijn Janssen, and P. Vigneswara Ilavarasan. 2019. "Perceived Usefulness, Ease of Use and User Acceptance of Blockchain Technology for Digital Transactions—Insights from User-Generated Content on Twitter." *Enterprise Information Systems* **13**, no. 6: 771–800. doi: [10.1080/17517575.2019.1599446](https://doi.org/10.1080/17517575.2019.1599446)
- Kretinin, Andrey, Jim Samuel, and Rajiv Kashyap. 2018. "When the Going Gets Tough, the Tweets Get Going! An Exploratory Analysis of Tweets Sentiments in the Stock Market." *American Journal of Management* **18**, no. 5: 23–36.
- Koufaris, Marios. 2002. "Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior." *Information Systems Research* **13**, no. 2: 205–23. doi: [10.1287/isre.13.2.205.83](https://doi.org/10.1287/isre.13.2.205.83)
- Lathrop, Scott. 2016. "A Call to Action to Prepare the High-Performance Computing Workforce." *Computing in Science & Engineering* **18**, no. 6: 80–3. doi: [10.1109/MCSE.2016.101](https://doi.org/10.1109/MCSE.2016.101)
- Moore, Gordon E. 1965. "Cramming More Components onto Integrated Circuits." *Proceedings of the IEEE* **86**, no. 1 (1998): 82–85.
- Moore, Gordon E. 1995. "Lithography and the Future of Moore's Law." In *Integrated Circuit Metrology, Inspection, and Process Control IX*, volume 2439, 2–17. International Society for Optics and Photonics.

- Moran, Nuala, and Joanne O’Dea. 2013. “Prace Special Report Supercomputers For All—The Next Frontier for High Performance Computing.” Brussels, Belgium: Science | Business Publishing Ltd. Accessed September 9, 2021, [https://www.gen-ci.fr/sites/default/files/prace\\_report\\_october\\_2013.pdf](https://www.gen-ci.fr/sites/default/files/prace_report_october_2013.pdf).
- Mosin, S. 2017. “The State of the Art Trends in Education Strategy for Sustainable Development of the High Performance Computing Ecosystem.” In *Russian Supercomputing Days*, 494–504. San Diego: Springer.
- Neeman, Henry, Hussein M. Al-Azzawi, Aaron Bergstrom, Zoe K. Braiterman, Dana Brunson, Dirk Colbry, Eduardo Colmenares, et al. 2018. “Progress Update on the Development and Implementation of the Advanced Cyberinfrastructure Research & Education Facilitators Virtual Residency Program.” *Proceedings of the Practice and Experience on Advanced Research Computing*, 1–7.
- Neeman, Henry, Aaron Bergstrom, Dana Brunson, Carrie Ganote, Zane Gray, Brian Guilfoos, Robert Kalescky, et al. 2016. “The Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency: Toward a National Cyberinfrastructure Workforce.” *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*, 1–8.
- Neumann, Philipp, Christoph Kowitz, Felix Schraner, and Dmitrii Azarnykh. 2017. “Interdisciplinary Teamwork in HPC Education: Challenges, Concepts, and Outcomes.” *Journal of Parallel and Distributed Computing* **105**: 83–91. doi: [10.1016/j.jpdc.2016.12.025](https://doi.org/10.1016/j.jpdc.2016.12.025)
- Obieniu, A. C., and F. I. Amadin. 2021. “User Acceptance of Learning Innovation: A Structural Equation Modelling Based on the GUAM Framework.” *Education and Information Technologies* **26**, no. 2: 2091–33. doi: [10.1007/s10639-020-10341-x](https://doi.org/10.1007/s10639-020-10341-x)
- Parashar, Manish. 2019. “CALIBURN: Advanced Cyberinfrastructure Report.” *Rutgers Discovery Informatics Institute Report*. Rutgers: The University of New Jersey.
- Puertas-Martín, Savíns, Antonio J. Banegas-Luna, María Paredes-Ramos, Juana L. Redondo, Pilar M. Ortigosa, Ol’ha O. Brovarets’, and Horacio Pérez-Sánchez. 2020. “Is High Performance Computing a Requirement for Novel Drug Discovery and How Will This Impact Academic Efforts?” *Expert Opinion on Drug Discovery* **15**, no. 9: 981–85. doi: [10.1080/17460441.2020.1758664](https://doi.org/10.1080/17460441.2020.1758664)
- Rahman, Md. Mokhlesur, G. G. Md. Nawaz Ali, Xue Jun Li, Kamal Chandra Paul, and Peter H. J. Chong. 2020. “Twitter and Census Data Analytics to Explore Socioeconomic Factors for Post-Covid-19 Reopening Sentiment.” <https://arxiv.org/pdf/2007.00054>
- Rietz, Tim, Ivo Benke, and Alexander Maedche. 2019. “The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance.” In *Proceedings of the 14th International Conference on Wirtschaftsinformatik*, Siegen, Germany, February 24–27.
- Samuel, Jim. 2016. “An Analysis of Technological Features Enabled Management of Information Facets.” PhD diss., CUNY Academic Works.
- Samuel, Jim. 2017. “Information Token Driven Machine Learning for Electronic Markets: Performance Effects in Behavioral Financial Big Data Analytics.” *Journal of Information Systems and Technology Management* **14**, no. 3: 371–83. doi: [10.4301/S1807-17752017000300005](https://doi.org/10.4301/S1807-17752017000300005)
- Samuel, Jim. 2021. “A Call for Proactive Policies for Informatics and Artificial Intelligence Technologies.” *SSN: Scholars.org*. Accessed December, 2021. <https://scholars.org/contribution/call-proactive-policies-informatics-and>.
- Samuel, Jim, Md. Rahman Gg Ali, Ek Esawi, Yana Samuel, et al. 2020a. “Covid-19 Public Sentiment Insights and Machine Learning for Tweets Classification.” *Information* **11**, no. 6: 314. doi: [10.3390/info11060314](https://doi.org/10.3390/info11060314)
- Samuel, Jim, Rajiv Kashyap, and Andrey Kretinin. 2018. “Going Where the Tweets Get Moving! An Explorative Analysis of Tweets Sentiments in the Stock Market.” *Proceedings of the Northeast Business & Economics Association*, NJ, USA
- Samuel, Jim, Rajiv Kashyap, Yana Samuel, and Alexander Pelaez. 2022. “Adaptive Cognitive Fit: Artificial Intelligence Augmented Management of Information Facets and Representations.” *International Journal of Information Management* **65**: 102505. doi: [10.1016/j.ijinfomgt.2022.102505](https://doi.org/10.1016/j.ijinfomgt.2022.102505)
- Samuel, Jim, Ratnakar Palle, and Eduardo Soares. 2021. “Textual Data Distributions: Kullback Leibler Textual Distributions Contrasts on GPT-2 Generated Texts with Supervised, Unsupervised Learning on Vaccine & Market Topics & Sentiment.” Preprint, submitted June, 2021. *SSRN*, <http://ssrn.com/abstract=3856396>.
- Samuel, Jim, and Alexander Pelaez. 2017. “Informatics in Information Richness: A Market Mover? An Examination of Information Richness in Electronic Markets.” *Proceedings of the 8th International Conference on Society and Information Technologies, ICSIT*, FL, USA
- Samuel, Jim, Md. Mokhlesur Rahman, G. G. Md. Nawaz Ali, Yana Samuel, Alexander Pelaez, Peter Han Joo Chong, and Michael Yakubov. 2020b. “Feeling Positive about Reopening? new Normal Scenarios from Covid-19 us Reopen Sentiment Analytics.” *IEEE Access* **8**: 142173–90. doi: [10.1109/ACCESS.2020.3013933](https://doi.org/10.1109/ACCESS.2020.3013933)
- Samuel, Yana, Jean George, and Jim Samuel. 2018. “Beyond STEM, How Can Women Engage Big Data, Analytics, Robotics and Artificial Intelligence? An Exploratory Analysis of Confidence and Educational Factors in the Emerging Technology Waves Influencing the Role of, and Impact Upon, Women.” In *2018 NEDSI Annual Conference (47th)* (p. 359)

- Sancho, Maria-Ribera. 2016. "BSC Best Practices in Professional Training and Teaching for the HPC Ecosystem." *Journal of Computational Science* **14**: 74–7. doi: [10.1016/j.jocs.2015.12.004](https://doi.org/10.1016/j.jocs.2015.12.004)
- Schaller, Robert R. 1997. "Moore's Law: Past, Present and Future." *IEEE Spectrum* **34**, no. 6: 52–9. doi: [10.1109/6.591665](https://doi.org/10.1109/6.591665)
- Theis, Thomas N., and H.-S. Philip Wong. 2017. "The End of Moore's Law: A New Beginning for Information Technology." *Computing in Science & Engineering* **19**, no. 2: 41–50. doi: [10.1109/MCSE.2017.29](https://doi.org/10.1109/MCSE.2017.29)
- Venkatesh, Viswanath, and Michael G. Morris. 2000. "Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior." *MIS Quarterly* **24**, no. 1: 115–39. doi: [10.2307/3250981](https://doi.org/10.2307/3250981)
- Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. "User Acceptance of Information Technology: Toward a Unified View." *MIS Quarterly* **27**, no. 3 425–78. doi: [10.2307/30036540](https://doi.org/10.2307/30036540)