Is Innovation Shaped by Masculine Norms? A Longitudinal Case Study of a Consumer Product

Caseysimone Ballestas*
Department of Mechanical Engineering
University of California, Berkeley
Berkeley, California, USA
caseysimone@berkeley.edu

Samantha Kang*
School of Mechanical, Industrial, and Manufacturing Engineering
Oregon State University
Corvallis, Oregon, USA
kangsa@oregonstate.edu

Aimee Huff
College of Business
Oregon State University
Corvalis, Oregon, USA
aimee.huff@oregonstate.edu

Andy Dong
Oregon State University
Corvallis, Oregon, USA
dongan@oregonstate.edu

Kosa Goucher-Lambert
Department of Mechanical Engineering
University of California
Berkeley, California, USA
kosa@berkeley.edu

Figure 1: Hyper-Masculine Colorways in the Official Apple Watch Bands

Abstract

In theories and metrics of product innovation, gender is invisible or ignored, and innovative products are presumed to be gender-neutral or agnostic. Yet, many ostensibly-innovative consumer products overlook the needs of women and gender non-conforming individuals, suggesting an implicit masculine framing. This research introduces a mixed-methods approach for analyzing gender scripts in product features and marketing, applied to a case study of the Apple Watch (2015–2024). Findings reveal a sustained reinforcement of gender norms: masculine-coded language and industrial design dominate how innovation is presented, even as objective technical improvements decline. In contrast, feminine-coded features, especially relational or user-centered ones, receive less emphasis in innovation framing. This work demonstrates how masculine value systems shape perceptions and theories of innovation and offers opportunities for future research on gender and design.

CCS Concepts

• **Human-centered computing → HCI theory, concepts and models**; **Interaction design theory, concepts and paradigms**.

Keywords

Innovation, Case Study, Gender, Product Assessment, Design Theory, Apple

ACM Reference Format:
Caseysimone Ballestas, Samantha Kang, Aimee Huff, Andy Dong, and Kosa Goucher-Lambert. 2025. Is Innovation Shaped by Masculine Norms? A Longitudinal Case Study of a Consumer Product. In *Designing Interactive Systems Conference (DIS '25), July 05–09, 2025, Funchal, Portugal.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3715336.3735834

1 Introduction

Interactive technologies support and enhance how people engage with both the digital and physical world, offering practical and meaningful user experiences [62]. Innovation in interaction design develops new interactive technologies and the capabilities these technologies can afford. Innovative designs enhance interaction modalities through novel functionality, optimized performance improvements, improved architecture, new external interactions or user interactions, and lower costs [56]. Product designers use various methods to achieve technological innovation and design newness [44, 62]; however, while product innovation may appear systematic and neutral, the process and its outcomes can reflect

*Caseysimone Ballestas and Samantha Kang contributed equally to this research and are co-corresponding authors.

DIS '25, Funchal, Portugal

ACM ISBN 979-8-4007-1485-6/25/07
https://doi.org/10.1145/3715336.3735834

and reinforce institutionalized knowledge, values, and action [18, 25, 30, 80].

Design scholars like Costanza-Chock [18] have critiqued design theory and practice for inadequately examining social norms like gender and its impact on technology development. Wajcman and other techno-feminist scholars have shown that the design and development of technology is fundamentally infused with patriarchal power [29, 61, 79, 80]. This suggests that technological innovation is a non-neutral process and is influenced by the social construct of gender [24, 35].

When the concept of innovation is treated, without critical evaluation, as a metric that is neutral *and* agnostic of gender, designers may inadvertently design products for masculine norms, even when their goal was to design for innovation [19]. It is therefore critical that product designers understand how the innovative technologies they create can reflect these social constructions of gender, and that scholars in the Human-Computer Interaction and Design (HCI&D) communities can conceptualize and theorize gender's role in innovation.

To address this problem, this work seeks to interrogate whether the dominant notion of innovation—that it is neutral and agnostic of other factors, specifically gender—holds. The literature not only suggests a connection between gender and interaction design [40, 65, 66, 81], but also indicates that the binary gender norms manifest in multiple dimensions of products [25, 30, 76, 79]. With this in mind, our primary research goal was to understand the construction of innovation across a relatively gender-neutral consumer product's design history and the extent to which that construction is co-constructed, reinforced, or perpetuated alongside gender norms. To do so, we asked the following questions:

RQ1 *How is innovation conceptualized and presented in an interactive consumer product's technical specifications and external documentation?*
RQ2 *How is gender conceptualized and presented in an interactive consumer product's industrial specifications and external documentation?*

In our study, we examine a ten-year (2015-2024) exemplar case study of an interactive consumer product. In addition to our empirical results, our research contributes (1) a novel codebook of gender scripts in industrial design specifications and external documentation and (2) a comprehensive longitudinal dataset that can be used in future research or educational settings to examine the construct of gender within technological innovations.

1.1 Terminology

Throughout this paper, when we refer to our focal product, the *'Apple Watch,'* we refer to a product group that comprises three distinct product *lines*: the *'Series,'* *'SE,'* and *'Ultra.'* When referencing a specific product line, we append the term to clarify scope (e.g., *'Apple Watch Series .'*) Each product line includes one or more *'models,'* which we define as distinct watch body designs introduced during a given annual release. We use the term *'release'* to refer to the annual product launch window during which one or more models may be introduced. For example, the Series 8, SE 2, and Ultra 1 are each separate models from three different product lines, all introduced in Release 8 (2022). Additionally, each model is comprised of multiple *'variants,'* which are the industrial design variations available for that model in that release year. For example, the Series 8 model comes in multiple color and material variants.

Figure 2: Comparison of masculine and feminine design aesthetics in fragrance packaging, used to illustrate the principles of *separation* and *hierarchy* within the gender value system (GVS) (see Sec. 2.3). Differences in form, material, and color reflect gendered cues aligned with binary norms: sharpness, size, and darkness signal masculinity; softness, ornamentation, and pastel palettes signal femininity. These distinctions reproduce normative gendered values and hierarchies in product design. From L to R: women's fragrances *Good Girl Blush* by Carolina Herrera [14] and *Dolce Garden* by Dolce & Gabbana [23] and men's fragrances *Uomo Intense* by Valentino [74] and *Spicebomb* by Viktor & Rolf [52]. Product Marketing Images: © Carolina Herrera, Dolce & Gabbana, Valentino, and Viktor & Rolf, respectively.

2 Related Work

Academic discussions of innovation and gender in the Human-Computer Interaction and Design (HCI&D) literature, as well as their nearby critical theory neighbors in feminist technoscience and feminist science and technology studies (STS) are not entirely new. In this section, we describe the history of normative framings of both concepts before reviewing how feminist and critical scholars have challenged these framings. Our aim is to introduce the theoretical foundation on which our work builds and to clarify where our contribution seeks to extend this critical tradition.

2.1 Innovation as a (normative) construct in HCI and Design

While definitions of 'innovation' vary, the literature consistently describes innovation in close relationship with ideas of progress [38, 57]. This conceptual linking of innovation and progress can be traced to early twentieth-century Western economic theories, particularly the work of Austrian economist Joseph Schumpeter (1883-1950) who conceived of innovation (i.e. physical/technological product) as a primary driver of economic development [59, 60]. In that context, economic development was effectively positioned as a proxy for progress [27]. This foundational logic which conflates innovation with progress (i.e., economic development) continues to inform theories about what innovation *is* and models that describe *how it works*.

Framed through this lens, dominant theories of innovation share a familiar set of qualities—qualities often treated as self-evident despite their ideology, including:

- **A persistent techno-optimism:** technological advancement is assumed to be both inevitable and desirable [11, 21];
- **A conflation of innovation with:** societal improvement, economic growth [28, 57] and enhanced individual performance [50];
- **A responsiveness to consumer 'demands' and 'desires':** markets serve as both the stimulus and the validation for innovation [3];
- **A preoccupation with novelty:** what is 'new,' 'better,' and above all, 'commercializable' is good [57].

However, these theoretical assumptions about the qualities that comprise innovation are not just rhetorical, they are embedded in the models that operationalize innovation in practice across the fields of design and engineering. For example, evolution-inspired models of innovation, such as the Technology Life Cycle model and S-Curve pattern of innovation, cast innovation as unfolding in predictable stages—starting with periods of continuous change, followed by discontinuous change in product performance over time [1, 41, 71]. In contrast, linear models of innovation, such as the Need-Pull and Technology Push-Demand Pull models, portray innovation as a driver of market development that responds to clearly defined consumer preferences [15, 31].

In HCI&D scholarship, reflexively taking a progress-oriented view of innovation continues to implicitly shape how the term is used and understood. As a result, innovation is often described as the development of new technologies, features, or processes which can shift the nature of the marketplace, address latent customer needs, or fulfill existing needs in a significantly new way [56, 62]—outcomes which are, normatively, treated as inherently good and desirable. This reflexively-assumed-framing: (1) makes it difficult to question the coherence of innovation or treat it as anything other than neutral or objective; (2) obscures the fact that progress itself is a contested and value-laden category; and (3) imports particular, often unexamined, assumptions about what counts as progress and, thus, innovation. Upon recognizing this entanglement, we can begin to ask critical questions about what kind of innovation-as-progress is being pursued, by whom, and for whose benefit.

Design scholars like Verganti [77], Chisholm [16], and Bucolo [12] have taken up this challenge, in part. They argue that while the kind of physical/technological product innovation initially theorized by Schumpeter is important, innovation also occurs at a conceptual level (such as through shifts in meaning, changes in value systems, and reconfigurations of user narratives). Their work, and others work like it, broadens the scope of innovation beyond changes in form or function alone, pointing to how innovation can also reshape the symbolic or experiential dimensions of design. However, while their work expands what innovation *does*, it stops short of interrogating the coherence and stability of what innovation *is* as a concept—including how the working definition of innovation is constructed through inherited ideas and values about progress. That is, it fails to examine how dominant ideas about innovation are themselves constructed in relation to normative assumptions about progress and that innovation is inherently objective, neutral, and good.

As widely influential as these models are, especially in the context of design and technology roadmapping and business development [37], critical theory scholars have long critiqued them for lacking deeper engagement with social and political factors that co-construct their value. Feminist and STS scholars, in particular, argue that because technology is created within a social framework, it necessarily has political implications, which, in turn, shape its ability to be either beneficial or harmful to different social groups [30, 38, 79]. While these critiques are not centered on interrogating the concept of innovation—rather, the outputs of a society that valorizes it—recent work by Kang & Dong [35] follows this thread further, interrogating innovation itself. Their critique of dominant models like the Technology Life Cycle hinges on the omission of powerful social constructs like gender as an explanatory variable in technology trajectories. Their empirical work on the design history of the sewing machine highlights the role of gender norms, especially around labor and domesticity, and their influence on the machine's technology development. Following their lead, this work builds on this critical turn by centering attention on the concept of innovation as a co-produced category—one that both shapes, and is shaped by, gender norms specifically.

2.2 Gender (as a normative construct) in HCI and Design

Notably, while definitions of 'gender' also vary, critical feminist scholarship consistently positions gender not as a fixed identity, but as a normative social system shaped by power, hierarchy, and cultural expectation. This conceptualization traces back to mid-20th century when Western feminist scholars who argued that the association between masculinity and technological control is historically constructed and politically maintained and that men's monopoly of technology is an important source of control and power [19, 29, 30, 33, 79]. Within this framework, gender operates as a symbolic and structural order that assigns value through binary oppositions—typically privileging the male/masculine over the female/feminine. This foundational logic continues to inform the fields of HCI&D, encoding and reproducing assumptions about who consumers are and what they 'demand' and 'desire' [6, 10].

Framed through this lens, dominant constructions of gender in design and technology tend to reflect several interlocking ideological qualities, including:

- **A persistent association of masculinity with**: rationality, technical mastery, and innovation [30, 46, 79];
- **A portrayal of femininity as**: emotional, aesthetic, or secondary to function [25, 75, 76, 80];
- **The centering of male bodies and experiences as**: defaults in design [10, 30];
- **The treatment of gender as**: static, binary, and visually legible [13, 30, 82].

Subsequently, the field of HCI&D is not value-neutral and neither technology nor innovation are neutral objectives since both the processes and artifacts created by and through these concepts and endeavors are male-normed and -centric [6, 10]. Rather, the field inherently embeds or co-produces values and norms around not only gender, but also race, class, and other identities in both its knowledge and artifact creation [10].

We build on this critical and feminist scholarship by adopting a social construction view of gender; that is, gender, its identification, expression, and performance [13] is not fixed and may or may not align with biological sex [32]. At the personal level, gender is something one *'does,'* rather than something inherent and intrinsic to who someone *'is'* [22, 82]. At the societal level, gender is something that is leveraged and maintained as a social construct in order to maintain power, organize labor, and restrict sexuality in patriarchal societies [79]. This "co-production," a term originating from Faulkner [30], of gender and theories about gender are poorly understood within the HCI&D community [10]. For this reason, classic exemplars of gender scripts embedded in technology are not widely known outside of the corners of discourse where scholars are explicitly discussing gender and other identities. Classic examples include voice recognition systems that unreliably recognize female voices [4, 49], airbags and car safety features designed using male physiques [54], iPhones that are too large for women's hands and snag long hair [72], and earlier versions of the iPhone health app that failed to track menstrual cycles until later updates [10]. More recent examples of gender scripts in HCI&D include Cryan [20], McKay [42], Offenwanger [45], and Otterbacher [47]. Cryan highlighted the lack of agreement on the qualification of gender stereotypes in rhetoric, especially in natural language processing, a limitation that is also relevant to this work [20]. McKay explored how gender and conference participation intersect, finding that attendance is shaped by factors such as sub-conference venues and location [42]. This is comparable to our attempt to investigate how gender and innovation intersect. Offenwanger identified the under-representation of women and non-binary individuals in study participation, noting that women are often intentionally recruited while men are recruited by coincidence [45]. Certain research topics, such as haptics and touch input—dimensions present in the technical specifications dataset for our case study—tend to have a more masculine or male-dominated representation in the study participant pool [45]. Otterbacher examined gender stereotypes in the search for images, finding that men are depicted more frequently than women, except when the search focuses on feminine traits such as warmth or emotionality (concepts that are repeatedly found to be gendered including in word inventories based on the BSRI and PAQ) [47].

Acknowledgment of the harms that are perpetuated when gender is not considered, especially for gender-non-normative users of HCI&D processes and artifacts has been relatively well documented in the literature. However, we identify two gaps in this robust literature, and we seek to address them in this work. First, there is a lack of analytical approaches to analyze, explain, and critically interpret the materialization of gender norms in consumer products. Second, more exemplars of dominant concepts in the field need to be interrogated for their possible co-production with gender. Instead, we see gender scripts and co-productions in artifacts critiqued (e.g., *air bags* or *phones*) especially, but less so in constructs or concepts (e.g., *innovation* or *participation* [42]) themselves.

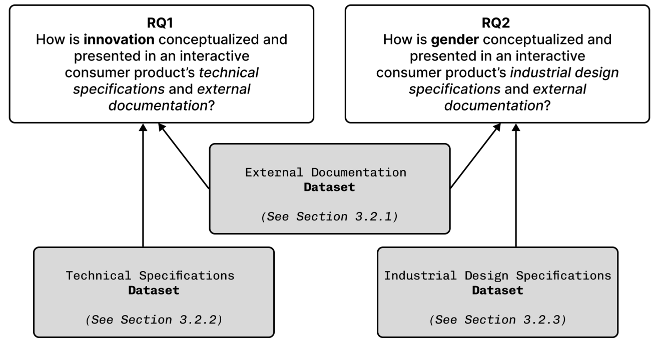


Figure 3: Flowchart showing how each of the three datasets map onto the study's two research questions about innovation (RQ1) and gender (RQ2).

2.3 Gender Value System and Metrics of Innovation

We ground our investigation on what gender researchers term the **gender value system (GVS)** (or gender system or gender order). The GVS is a power structure (norm) that organizes the relationship between sexes on a symbolic, structural, and individual level [2, 34]. The system is built according to principles of separation and hierarchy between genders. The *principle of hierarchy* theorizes that the male-masculine being is the true standard in society and superior to that of the feminized being. The *principle of separation* theorizes that behaviors and tasks are divided into a binary norm, consisting of two separate domains: the male-masculine and female-feminine domains.

The GVS's principles manifest in design in numerous ways as demonstrated by the scholarship of Ehrnberger *et al.* [25] and Moss [43]. For example, the principle of separation is often illustrated walking down the personal hygiene aisle at a pharmacy. Products designed for men, whether deodorant, fragrance, or razors, are characterized by masculine form factors such as machine aesthetics, straight angular and aggressive lines, large sizes, a focus on functional and technical features, complex gadgetry, dark colors, high performance, and instills a sense of danger, adventure, or challenge [25, 43]. In contrast, products designed to appeal to women tend to be characterized by feminine form factors like soft surfaces, round and organic shapes and lines, smaller size, pastel or bright colors (pink), and aesthetic and non-functional orientation [25, 43]. To illustrate, perfumes as shown in Fig. 2 seem strongly aligned with these masculine and feminine form factors and colorways.

3 Method

To attend to our research goal of understanding the construction of innovation across a focal consumer product's design history and the extent to which that construction is co-constructed, reinforced, or perpetuated alongside gender norms, we developed a multi-phase, mixed-methods approach centered on a longitudinal case study analysis that is systematically coded using a novel codebook allowing for content analysis. Our approach builds on the conceptual theoretical lenses outlined in the background, operationalizing these ideas and the Gender Value System (GVS) across three interrelated datasets (See Fig. 3) which represent the attributes of inquiry of our focal product.

3.1 Case Study Criteria and Selection

We identified five criteria that the interactive consumer product needed to fulfill in order to be appropriate for a longitudinal case study:

(1) **Physical Tangibility and Computational.** The product needed to be physically tangible, specifically as a hardware-software system, rather than existing purely as a service or digital offering. This material presence allows for a grounded analysis of how physical and digital design choices interact.
(2) **Longitudinal Continuity in Iteration.** The product needed to exhibit consistent, discrete design changes over a significant timeline. This temporal-continuity allows for tracking and analyzing the evolution of product changes over time, which would be less feasible with products lacking substantive or frequent iterative updates (i.e., limited-run items).
(3) **Consumer-Oriented Positioning.** The product needed to be consumer-facing, as opposed to a business-to-business (B2B) or industrial product. This consumer product orientation allows us to analyze dominant cultural and social norms reflected in product design choices.
(4) **Gender-Neutral Positioning.** The product needed to be perceived as broadly gender-neutral in its positioning. This neutrality allows for an examination of how gender norms might still be encoded despite a product's positioning.
(5) **Strategic Positioning as Innovative.** The product needed to come from a brand that positions itself as innovative. This strategic framing allows for the exploration of dominant narratives of progress and technological advancement.

Based on the five identified criteria, we considered several interactive consumer products before selecting the Apple Watch product group for our case study for meeting our five criteria in the following ways:

(1) **Physical Tangibility and Computational.** The Apple Watch is a hardware-software product that exemplifies tangible interaction. This tangibility and embedded computation support a grounded analysis of both technical *and* industrial design choices, aligning with both the interdisciplinary expertise of the researchers and the DIS community's focus on materiality and interfaces [51, 55].
(2) **Longitudinal Continuity in Iteration.** The Apple Watch product group has exhibited consistent, annual design iterations over a ten-year period (2015–2024), encompassing fifteen discrete model releases across three product lines. This temporal continuity offers the opportunity to track changes across significant time at scale, fulfilling our criterion for a longitudinal case study. Moreover, because this product has been worked on by many designers and engineers over the case study time period, it represents sustained institutional values perpetuated by a collective rather than any single authorial vision.
(3) **Consumer-Oriented Positioning.** The Apple Watch is explicitly designed, marketed, and sold as a consumer-facing product. Unlike a product designed for daily use, like a chair, the Apple Watch is a non-essential luxury consumer product, which makes it useful to examine how Apple aligns their marketing with dominant cultural norms around innovation and consumer desire.
(4) **Gender-Neutral Positioning.** The Apple Watch is consistently positioned as a gender-neutral product, marketed through imagery of people of all genders and framed as a customizable, visible personal technology. This makes it a particularly rich site for analyzing how gendered assumptions are subtly encoded in design and marketing.
(5) **Strategic Positioning as Innovative.** The Apple Watch is consistently positioned in the literature as an innovative product [69, 70]. Additionally, Apple as a brand has cultivated a long-standing reputation for innovative technology and design and technological leadership [70, 73, 78]. This association with innovation makes it of particular relevance for exploring how narratives of progress intersect with societal trends.

Additionally, though not part of our selection criteria, the Apple Watch holds the largest market share in the wearable device category, with more than 115 million watches sold since 2015 [67].

3.2 Dataset Construction

We constructed three datasets as part of this study (See Fig. 3). The first focused on *External Documentation* (See Sec. 3.2.1). The second focused on *Technical Specifications* (See Sec. 3.2.2). The third focused on *Industrial Design Specifications* (See Sec. 3.2.3). Our rationale for collecting these three datasets is that, together, they provide a comprehensive portrayal of our focal product, capturing its dimensions through engineering-design specifications *and* market positioning. All datasets spanned the entire range of examination (2015 to 2024) and included all three product lines (Series, Ultra, and SE) of the focal product group (Apple Watch).

Undergraduate research assistants (RAs) were recruited for this task. Dataset construction was advised and reviewed by the first two authors periodically for accuracy and consistency.

3.2.1 ***External Documentation Dataset.*** In this study, we were interested in collecting external documentation, by which we mean: public-facing corporate reports, memos, print ads, or other public-facing materials created by the brand of interest about the focal product. Given our chosen focal product was designed and engineered within a Californian-American cultural context, we concentrated specifically on its U.S.-based web presence, particularly its landing pages, as these represent a comprehensive repository of the product group's external communications. From these pages, we extracted the plain text and separated the data based on its HTML tags to prepare for coding and analysis. This process resulted in 16 web-pages (landing pages for all three product line's releases) totaling 36,824 words of text.

This dataset of plain text and expanded inventories served as a complementary mode of analysis for evaluating how Apple positioned gender and innovation through its marketing copy. We reasoned that the semantic content of each watch's flagship U.S. landing page provided a strong representation of what Apple intended consumers to perceive as the core rhetorical positioning of the product. In tandem with our two other datasets, this aligned

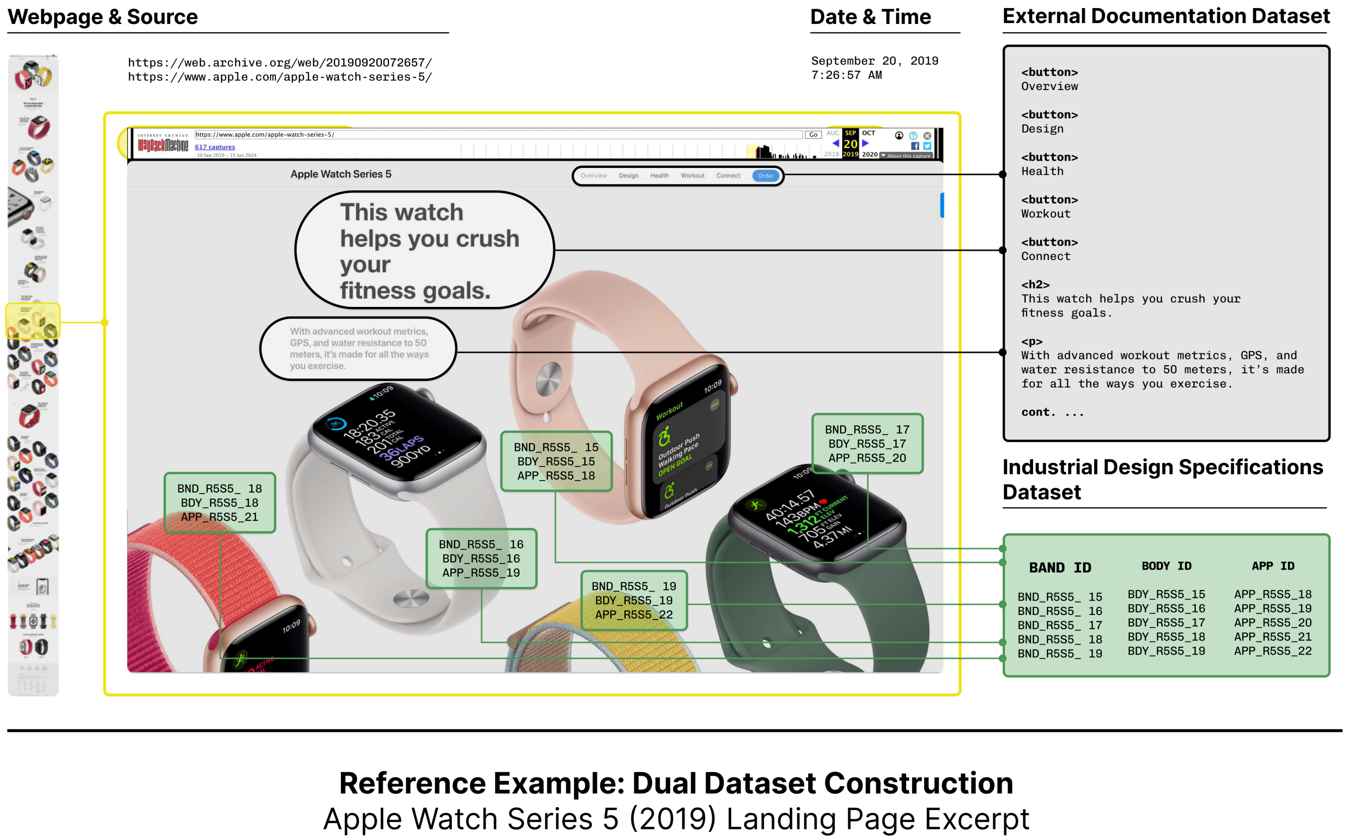


Figure 4: Annotated example of dual dataset construction from the Apple Watch Series 5 (2019) landing page. CENTER: Screenshot excerpt from the Wayback Machine shows five watch variants with labeled band, body, and app IDs contributing to the *Industrial Design Specifications Dataset* (see Sec. 3.2.3). RIGHT: HTML-extracted text and button elements structured into the *External Documentation Dataset* (see Sec. 3.2.1). TOP: Archival timestamp shows the first available Wayback Machine capture post-release. LEFT: Full-page snapshot with yellow box marking the excerpted region. Webpage Image Source: © Apple Inc.

with our aim of conducting a normative evaluation of gender and innovation through the lens of external documentation.

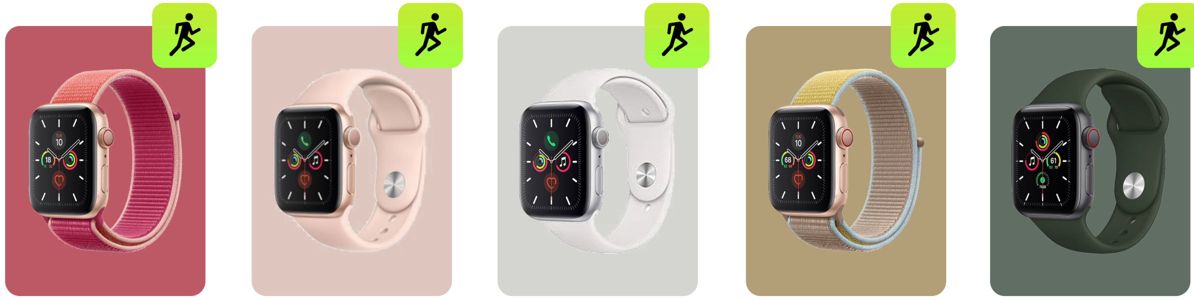
3.2.2 *Technical Specifications Dataset*.

We sourced information on eight dimensions (see below) of technical specifications primarily from Wikipedia, supplemented and verified with official Apple Watch specifications on Apple.com where possible using the Wayback Machine.

The *Technical Specifications Dataset* included the following attributes across eight dimensions:

(1) **System on Chip**.
(2) **Battery Life:** Battery life (hours), charge time (hours), and large/small body capacity (mAh, V, Wh).
(3) **Resistance:** Dust resistance (presence/absence) and water resistance (ISO number).
(4) **Chipset:** Central processing unit (bits, cores), random access memory (DRAM), and storage (GB).
(5) **Software:** System on chip (name), operating system (version), and phone requirement (model and iOS).
(6) **Input and Display:** "Siri Speaks" and "Raise to Speak" functionality (presence/absence), brightness (nits), display (type), Force Touch (presence/absence), pixel density (ppi), and screen dimensions (mm).
(7) **Sensors:** Accelerometer (g's), altimeter (presence/absence), ambient light sensor (presence/absence), blood oxygen sensor (presence/absence), compass (presence/absence), electrical heart sensor (ECG/EKG) (presence/absence), gyroscope (range), optical heart sensor (generation), and satellite navigation (type/s), temperature sensor (presence/absence).
(8) **Connectivity:** Bluetooth (version), cellular (type), ultra-wideband (chip type), and wireless networking(b/g/n and GHz).

This dataset served as a foundation for evaluating innovation within the Apple Watch product line and was the data source for the 'Innovation Scores' (See Sec. 3.6.2 for how score was calculated). We reasoned that the features highlighted as technical specifications, first on Apple.com and then captured by Wikipedia, provided



Watch Body	BDY_ID	BDY_R5S5_18		BDY_R5S5_15		BDY_R5S5_16		BDY_R5S5_20		BDY_R5S5_17	
	Material	Aluminum	+ 1.0	Aluminum	+ 1.0	Aluminum	+ 1.0	Aluminum	+ 1.0	Aluminum	+ 1.0
	Color	Gold	- 1.0	Gold	- 1.0	Silver	+ 1.0	Gold	- 1.0	Space Gray	+ 2.0
	Hex Code	#e9d198	—	#e9d198	—	#3f3d42	—	#e9d198	—	#c0c0c0	—
	Size	40 mm & 44 mm	+ 1.0	40 mm & 44 mm	+ 1.0	40 mm & 44 mm	+ 1.0	40 mm & 44 mm	+ 1.0	40 mm & 44 mm	+ 1.0
	Collab	No	0.0	No	0.0	No	0.0	No	0.0	No	0.0
Apps	APP_ID	APP_R5S5_21		APP_R5S5_18		APP_R5S5_19		APP_R5S5_23		APP_R5S5_20	
	Category	Wellness	—	Wellness	-	Wellness	—	Wellness	—	Wellness	—
	App	Workout	+ 2.0	Workout	+ 2.0	Workout	+ 2.0	Workout	+ 2.0	Workout	+ 2.0
	Action	Log	—	Goals	—	Log	—	Route	—	Log	—
Watch Band	BND_ID	BND_R5S5_ 18		BND_R5S5_ 15		BND_R5S5_ 16		BND_R5S5_ 20		BND_R5S5_17	
	Material/ Style/ Clasp	Woven Nylon/ Sport Loop/ Woven Nylon	+ 1.0	Fluoroelastomer/ Sport Band/ Aluminum	+ 1.0	Fluoroelastomer/ Sport Band/ Ceramic	+ 1.0	Woven Nylon/ Sport Loop/ Woven Nylon	+ 1.0	Fluoroelastomer/ Sport Band/ Aluminum	+ 1.0
	Color	Pomegranate	- 2.0	Pink Sand	- 1.0	Soft White	- 1.0	Camel	+ 1.0	Pine Green	+ 1.0
	Hex Code	#bc5864	—	#d7c1c1	—	#ebeae8	—	#b19f77	—	#606e64	—
	Collab	No	0.0	No	0.0	No	0.0	No	0.0	No	0.0
	GENDER SCORE	Feminine	- 0.25		0.00	Masculine	+ 0.8		+ 1.0	Hyper-Masculine	+ 1.25

Reference Example: Industrial Design Specification Dataset Gender Coding

Apple Watch Series 5 (2019) Models from Landing Page Snippet Excerpt

Figure 5: Breakdown of Gender Score assignment for five Apple Watch Series 5 (2019) models, based on visible combinations of watch body, band, and app from the landing page (see Fig.4). Each element was coded using a four-point binary framework: Hyper Feminine (–2), Feminine (–1), Masculine (+1), and Hyper Masculine (+2). These values were averaged to calculate a composite Gender Score per model, which was then mapped to categorical labels using the score ranges defined in Fig.7 (see also, Sec.3.5). Dashes (–) indicate elements that were collected but not assigned a gender code. Band IDs were cross-referenced using Bandbreite[5] for visual verification. The full set of coded variations and scores can be found in the extended dataset.

a strong representation of what Apple intended consumers to perceive as the core elements of the product. This aligned with our aim of conducting a normative evaluation of innovation through the lens of our focal product's technical specifications.

3.2.3 ***Industrial Design Specifications Dataset***. We sourced information on three dimensions (see below) from each line's landing page using the Wayback Machine. From there, we documented all visible Apple Watch lines. This process resulted in a dataset comprising 673 watches across 16 landing pages. Thereafter, we assigned values to features across our three dimensions of inquiry. During value assignment, we cross-referenced the data collected from Apple's official sites on watch bands with an application called Bandbreite, an archive of all official Apple Watch band styles and colorways [5]. This allowed us to assign values consistently for watch bands when Apple.com did not list design specification details on a given page.

The *Industrial Design Specifications Dataset* included the following attributes and variations across three dimensions:

(1) **Watch Body:** materials, sizes, colors, and brand collaborators.
(2) **Watch Band:** materials, style, colors, and brand collaborators.
(3) **Applications:** application function, application name, and application behavior.

This dataset served as a foundation for evaluating gender within the Apple Watch product line. We reasoned that the designs highlighted as landing-page-worthy across the lines provided a strong representation of what Apple intended consumers to perceive as

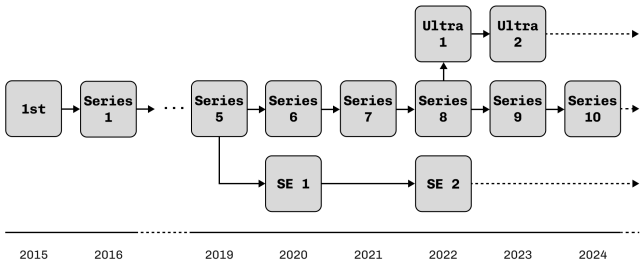


Figure 6: Timeline-based flowchart illustrating the release chronology and product line stratification of the Apple Watch from 2015 to 2024. The diagram displays three product lines—Series (middle row), SE (bottom row), and Ultra (top row)—with arrows indicating release sequences and ongoing production. Vertical connections show which models were used as reference points in our scoring methodology. For example, SE 1 was scored against Series 5 (its direct predecessor), while Ultra 1 was scored in relation to Series 8 (released concurrently). This stratified flow informed how we assigned comparative Innovation and Gender Scores across the product group.

the core design of the product. This aligned with our aim of conducting a normative evaluation of gender through the lens of tangible design elements.

3.3 External Documentation Codebook Development

We aggregated three existing word inventories (i.e., pre-compiled lists of items that describe particular traits or characteristics relevant to a particular term or concept) for 'masculine,' 'feminine,' and 'innovative.' The masculine and feminine inventories were drawn from foundational psychological instruments such as the Bem Sex Role Inventory (BSRI) [8] and the Personal Attributes Questionnaire (PAQ) [64], as well as from more recent HCI&D studies [20, 47, 58]. Innovation-related terms, by contrast, were drawn from a combination of trade publications [63] and the business economics academic literature [68]. These inventories were later expanded using natural language processing (NLP) techniques to identify synonyms and related terms in the Apple Watch corpus (e.g., 'advanced,' became: 'advance,' 'advancing,' 'advances,' etc.) and served as our codebook for scoring plain text extracted from the external documentation (see Sec. 3.6.1).

3.4 Technical Specifications Codebook Development

In developing our codebook for technical specifications, we drew from the normative accounts of innovation outlined in our background (see Sec.2.1). The frameworks discussed therein tend toward framing innovation as binary between what is considered progress and what is not. Accordingly, we assigned a score of +1 to any attribute (see Sec. 3.2.2) that demonstrated advancement relative to the previous model (e.g., improved battery life, added sensors, enhanced screen resolution), and a score of −1 to features that remained unchanged (stagnant) or were removed entirely (e.g., the discontinuation of Force Touch in the SE 2). We chose not to assign a neutral value (e.g., 0) to stagnant features, as the normative framing of innovation within design and engineering rarely considers stagnation to be innovative. Thus, the codebook reflects an operational binary: innovative (+1) versus not innovative (−1), in alignment with dominant assumptions about what constitutes progress.

3.5 Industrial Design Specifications Codebook Development

We developed a codebook to identify how gender is encoded in the focal product's industrial design specifications. The codebook was developed iteratively in a process that created a theoretically and practically robust Attribute x Code table which identified the gender code (and associated numerical score) associated with each variation of a discrete list of watch attributes. This enabled us to code using a systematic, objective approach akin to a visual content analysis as described by Rose [53] and Bell [7]. See Appendix A.

3.5.1 ***Phase 1: Initial Development.*** After discussing their positionalities as researchers, the authors collaborated over several weeks to develop the codebook for analyzing gender scripts in Apple Watch industrial design. This was an iterative process aimed to achieve theoretical grounding, consistent interpretation, and reliable results. Authors drafted initial codes based on established visual analysis methods, design and gender literature, and familiarity with the focal product. We identified specific attributes to be coded: band color (proprietary name and hex code), materials (proprietary name), collaboration (name), and style (proprietary name); body color (proprietary name and hex code), material (proprietary name), collaboration (name), and size (mm); and application (type, name, behavior). Preliminary codes were proposed and discussed in relation to a subset of images. Phase 1 codes included: Masculine (+1), Neutral (0), and Feminine (-1).

3.5.2 ***Phase 2: Refinement.*** Authors independently coded a subset of images and compared results. We discussed discrepancies, refined definitions, and added orienting language for each code category based on literature and our experience coding the subset. We revisited the literature and methods guides to refine the code book. In this phase, we added Hyper-Masculine (+2) and Hyper-Feminine (-2) codes.

3.5.3 ***Phase 3: Creating Attribute x Code Table.*** Undergraduate RAs, hereafter 'coders,' were recruited and made aware of the general topic of the study (gender in product design) but blind to the authors' specific propositions. Authors directed the coders to identify every variation of attributes in the *Industrial Design Specifications Dataset* and to create a discrete list of attribute variations to be coded. Authors then assigned a gender code to each variation. For example, the coders identified 28 unique watch band materials in the *Industrial Design Specifications Dataset*, and the authors assigned one gender code to each band material.

3.5.4 ***Phase 4: Training Coders.*** Using a subset of images of variants of the focal product, authors instructed the coders to assign the appropriate code based on the attribute variant in each image.

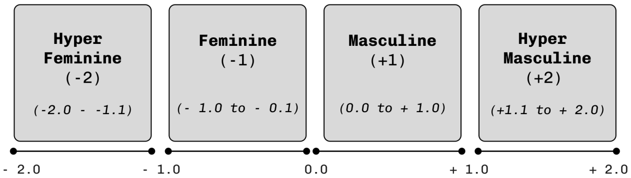


Figure 7: Mapping of Gender Score ranges to categorical codes used in analysis. While our coding schema applied discrete values to product elements—Hyper Feminine (-2), Feminine (-1), Masculine (+1), and Hyper Masculine (+2)—final Gender Scores reflect averaged values across multiple elements creating continuous values. This figure defines how the score ranges correspond to each category, allowing continuous values to be categorized within the four-point binary framework.

In this way, coders work was similar to a content analysis, in that their discretion and judgment were not involved in coding; they simply identified the attributes and applied the appropriate codes.

3.5.5 ***Phase 5: Final Revisions.*** The coded subset of images was reviewed by the authors for consistency, accuracy, and theoretical grounding. Edge cases and problematic codes were identified and discussed by the coders and authors. The Neutral (0) gender code was removed during this phase because, at the attribute level, we observed gendered value systems at play and could not establish construct validity.

3.6 Dataset Coding and Scoring

The coders coded the full dataset. The researchers and coders held regular calibration meetings and conducted spot checks for consistency. The authors and a separate RA who had not been involved in coding analyzed the data.

3.6.1 ***External Documentation's Gender and Innovation Scores.*** We treated the compiled and NLP-expanded word inventories for masculinity, femininity, and innovation as codebooks for content analysis. For each Apple Watch landing page, we counted all instances of terms from the masculine, feminine, and innovation inventories, then normalized these raw frequencies by the total word count of the page. This produced three proportional scores per page: a Masculine-Term Score, a Feminine-Term Score, and an Innovative-Term Score each representing the percentage of page text composed of terms rhetorically associated with that construct.

3.6.2 ***Technical Specifications' Innovation Score.*** We developed a logic for coding innovation by evaluating each model's features based on their newness or improvement. Features that were stagnant or regressive compared to prior models were assigned a score of -1, while features that were new or improved were assigned a score of +1. These scores were then averaged for each model to produce an overall Innovation Score.

Additionally, our approach incorporated a relational schema to account for the hierarchical nature of the tiered offerings across all three product lines. Series line models were compared to their Series predecessor. SE line models were first compared to their predecessor Series model (starting with Series 5) and subsequently to their predecessor SE model. This approach was grounded in the SE's positioning as a budget option within the product line. Ultra line models were first compared to their contemporaneously released Series model (e.g., Ultra to Series 8) and later to their predecessor Ultra model, reflecting its positioning as a higher-end option.

3.6.3 ***Industrial Design Specifications' Gender Score.*** Based on our theoretical approach informed by the Gender Value System (see Background 2), each watch attribute was coded as one of four gendered categories, and a numeric value was associated with each code: Hyper-Masculine (+2), Masculine (+1), Feminine (-1), or Hyper-Feminine (-2). See appendix for full codebook**??**. For each product line, the models' (band, body, and application) respective gender codes were compiled to produce an overall Gender Score. The final Gender Score for each model was calculated by averaging the scores for all coded attributes.

4 Results

We sought to critically evaluate how innovation and gender are conceptualized and operationalized within the Apple Watch product line, with the goal of revealing the normative assumptions embedded in both constructs. For innovation, we leveraged dominant theories, as outlined in Sec. 2.1, to score each model's technical specifications as either progressive (+1) or regressive/stagnant (-1), assigning a cumulative Innovation Score detailed in the methodology (see Sec. 3.6.2). For gender, we operationalized critical and feminist theory, as outlined in Sec. 2.2, which we translated into a codebook to evaluate industrial design specifications and external documentation, assigning a cumulative Gender Score detailed in the methodology (see Sec. 3.6.3). These frameworks allowed us to analyze the Apple Watch product line through consistent normative lenses, revealing normative trends in both innovation and gender over time.

Overall, across yearly averages, we observe a strong positive correlation between Gender Scores and Innovation Scores (r = +0.90), meaning that in years where the average Gender Score across all released models was coded as more Masculine, the average Innovation Score was also higher. This relationship holds across all Gender Score variants: Alternate A (bodies and bands only) shows the highest correlation with Innovation Scores (r = +0.96), while Alternate B (bodies, bands, and apps excluding watch faces) yields a similar correlation (r = +0.90). Rather than indicating opposition between gendered design and innovation, these results suggest alignment: design features that contribute to masculine coding may also be those most likely to register as innovative within this product category. In the following two subsections we dive deeper into our data.

4.1 How Innovation is Conceptualized and Presented

In response to our first RQ–(*How is innovation conceptualized and presented in a consumer interactive product's technical specifications and external documentation?*)–we identified two key findings.

First, our analysis reveals the challenge in sustaining a normative model of innovation—defined here as a continuous *and*

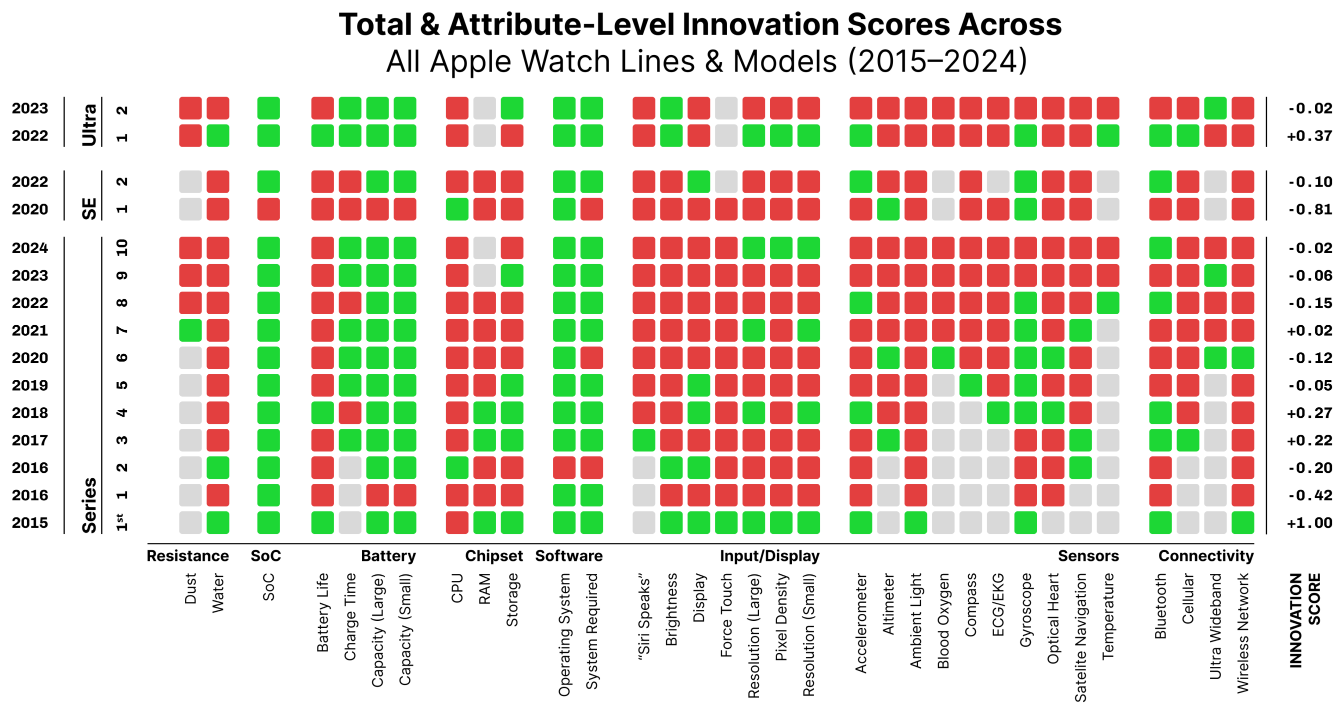


Figure 8: Each cell represents a binary-coded assessment of a specific attribute for a given Apple Watch model. Green (+1) indicates improvement over the previous model (e.g., longer battery life or better resolution); red (−1) indicates stagnation or removal of a feature; gray indicates the attribute was not applicable, absent, or couldn't be triangulated in dataset building. The Innovation Score at the end of each row reflects the average of all coded attributes for that model.

year-over-year introduction of new or improved technical features—i.e., innovation is not uniformly achieved across the Apple Watch product group. As shown in Fig. 8, Innovation Scores fluctuate considerably over the ten-year span. While the original release (Series 1st) scored highest, subsequent models in the Series line often reflected stagnation or even regression. Even when the initial launch is excluded as an outlier, patterns of incrementalism emerge: some years exhibit marked innovation (e.g., Series 3 and 7), while others show negligible gains (e.g., Series 4 and 10).

Second, innovation is not evenly distributed across models but is instead differentially allocated across product lines. The introduction of the SE (2020) and Ultra (2022) models exemplifies this approach. In each of these release years, the new model (SE 1 in 2020, Ultra 1 in 2022) scored lowest or highest, respectively, indicating product group feature stratification.

Third, innovation is weakly coupled with linguistic signaling. The correlation between technical Innovation Scores and the Innovative-Terms Score in external documentation is moderate overall ($r = 0.64$), but inconsistent across product lines. For example, SE models exhibit some of the lowest Innovation Scores in the dataset (e.g., −0.81 for SE 1), yet their use of innovative terms (e.g., Innovative Terms Score: 0.01%) is comparable to higher-scoring models like Ultra 1 (Innovation Score: +0.37; Innovative Terms Score: 0.02%). This suggests that the rhetorical signaling of innovation (via terms like "new" or "advanced") does not reliably correspond to substantive change in technical specifications.

Fourth, innovation is concentrated in a narrow band of feature categories. As Fig. 8 shows, most positive Innovation Scores cluster around a few specific attributes: battery life, display brightness, and the addition of new sensors (e.g., ECG, blood oxygen, temperature). Conversely, other areas such as chipsets, resolution, or connectivity technologies (e.g., Bluetooth, Wi-Fi) remain largely unchanged or are downgraded over time. This uneven distribution of improvements reflects how certain technical attributes have seen repeated advancement over time, while others remain unchanged—possibly due to the difficulty of improving specific features or im- or ex-plicit prioritization of particular dimensions of technical performance over others.

4.2 How Gender is Conceptualized and Presented

In response to our second RQ—(*How is gender conceptualized and presented in a consumer interactive product's industrial specifications and external documentation?*)—we identified three key findings.

First, Gender Scores trend increasingly Masculine over time. Across all three measurement variants (see Fig. 10), we observed an upward trajectory in Gender Scores across the Apple

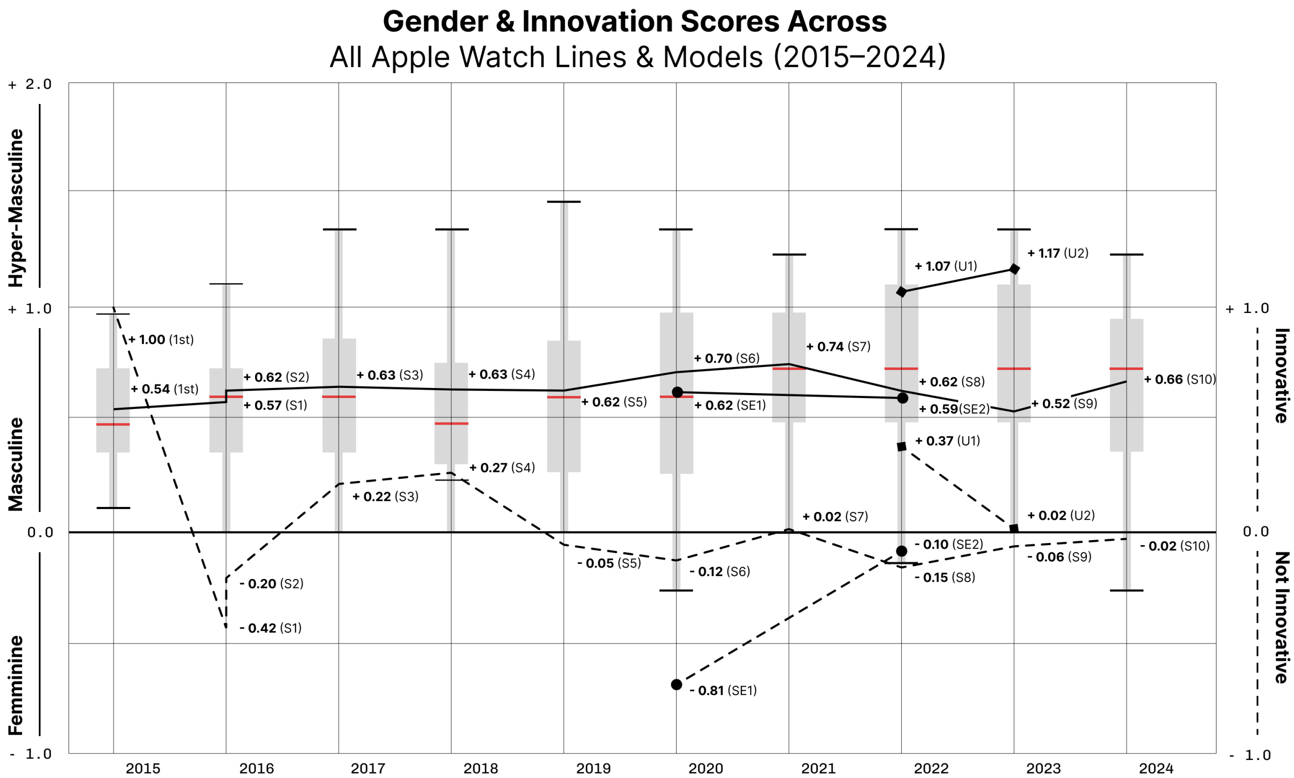


Figure 9: Comparison of Gender and Innovation Scores across Apple Watch models from 2015 to 2024, spanning the Series, SE, and Ultra lines. Solid black lines trace the average Gender Score for each model line, with shape-coded points: squares (Series), circles (SE), and diamonds (Ultra). Dashed black lines trace the corresponding Innovation Scores over time, with the same point styles. Box plots show the distribution of Gender Scores for each model release, including medians (red bars), interquartile ranges (gray boxes), and full range of scores (whiskers). Gender Scores are plotted on the left y-axis, from Feminine (−1) to Hyper-Masculine (+2); Innovation Scores are plotted on the right y-axis, from Not Innovative (−1) to Innovative (+1). Labels next to each point indicate model name and score. This figure illustrates the relative consistency of masculine-coded design across all lines, with Ultra models exhibiting the highest average Gender and Innovation Scores, SE models the lowest, and Series models falling in the middle.

Watch Series line, indicating a shift toward increasingly Masculine-coded design elements over time. These variants included: (1) *Alternate A*—scores based on industrial design attributes alone, i.e., bodies and bands (dashed red line), (2) *Alternate B*—scores based on bodies, bands, and applications *excluding* watch faces (dashed yellow line), and (3) the *Gender Score*—scores based on bodies, bands, and *all* applications, including watch faces (solid blue line). The Gender Score was selected for subsequent analysis because it captures the most comprehensive set of product elements. In the Series line there is a directional trend: Series models become more Masculine-coded over time, with variation depending on which features are included (see Fig. 10). Alternate A produces the weakest trend (R^2 = 0.091), reflecting the cumulative effect of only physical design attributes. Alternate B is slightly lower still (R^2 = 0.086), suggesting that excluding watch faces may dampen the overall trajectory. The Gender Score, our official measure, shows the strongest increase (R^2 = 0.119), with scores reaching into the Hyper-Masculine range (+1.0 to +2.0) in later models.

In the SE line, we see relatively consistent scores across both releases (SE1: +0.62, SE2: +0.59), each landing in the mid-Masculine range, with minor variation year-to-year (see Fig. 9). In the Ultra line, we see significantly higher Gender Scores from both the Series and SE, with both Ultra 1 and Ultra 2 scoring above +1.0 (U1: +1.07, U2: +1.17), placing them firmly in the Hyper-Masculine category (see Fig. 9). This upward slope across the Series and Ultra models and Masculine range for all lines, underscores how gendered coding intensifies over time when all product dimensions are considered, including software-based functionalities.

Second, no model acquired a Feminine- or Hyper-Feminine Gender Score. Despite instances of Feminine- and Hyper-Feminine coded variations in materials and colorways. Despite several band and body variations using Feminine-coded colorways—e.g., 'White' (a bright white with a rough hex value of #ffffff), 'Starlight' (a light cream with a rough hex value of #e4e2de) and 'Gold' (a gold or rose-gold-like metallic color in both matte and shiny color with rough hex values in the range from #deb9A9 #f1d8cc #ad8f7a, #dbc5a6,

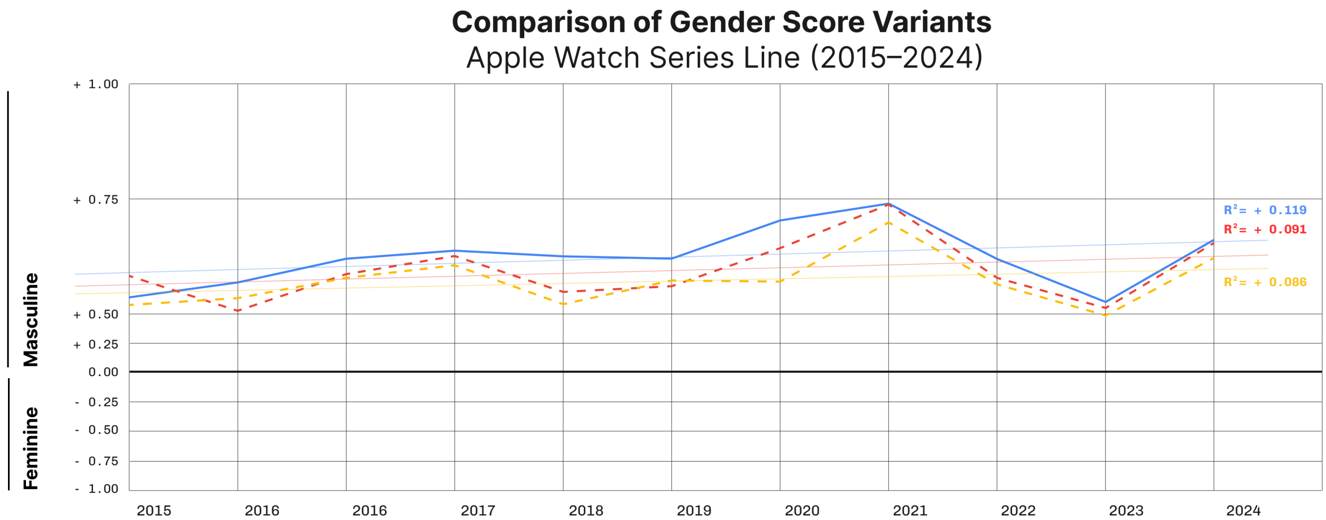


Figure 10: Comparison of three Gender Score variants across Apple Watch Series models from 2015 to 2024. The solid blue line represents the primary Gender Score, calculated by averaging gender-coded values for all bands, bodies, and applications per model release. Alternate A (dashed red line) includes only bands and bodies, omitting applications. Alternate B (dashed yellow line) includes applications but excludes lock screen apps (i.e., watch faces). Linear trendlines are overlaid in matching colors with corresponding R^2 values: +0.119 (primary), +0.091 (Alternate A), and +0.086 (Alternate B). All three variants show a general upward trend in masculinity, with the primary score selected for analysis due to its inclusion of both all digital and industrial design elements.

#ebdbc3)—these features were often offset by Masculine- or Hyper-Masculine coded design elements in size, material, or color (e.g., 'Black' with a rough hex code of #000000, 'Midnight' with a rough hex code of #44464c, 'Graphite' with a rough hex code of #45433f), especially in the Series and Ultra lines. The absence of gynocentric sizing and the inconsistent use of Feminine- and Hyper-Feminine coded bands limited the influence of these individual Feminine- and Hyper-Feminine coded design elements on the final Gender Scores. As a result, all models were ultimately coded as Masculine or Hyper-Masculine.

The vertical box plots in Fig. 9 offer additional insight to the Masculine and Hyper-Masculine distribution of Gender Scores across all industrial design variations released in each year. Across the ten-year span, every model year showed a median Gender Score (red line) in the Masculine or Hyper-Masculine range, with several years, particularly 2022 and 2023, showing a wide interquartile range extending well above +1.0. Notably, no year shows a median or lower quartile extending into the Feminine range.

Third, gendered language in external documentation reveals a subtle but consistent semantic trend over time. While gendered terms appear infrequently overall, patterns across the product lines suggest a shift in lexical framing. In the Series line, Feminine-coded terms declined from 0.013% in the 1st model to 0.000% by Series 10, while Masculine-coded terms fluctuated but increased slightly overall (e.g., from 0.0104% to 0.0200%). This directional trend holds for the SE and Ultra lines as well, where Masculine terms consistently outnumber Feminine ones (e.g., SE 1: 0.0063% M vs. 0.0005% F; Ultra 1: 0.0048% M vs. 0.0003% F). Correlation analysis affirms these associations. Across the full dataset, Gender Scores negatively correlate with the proportion of Feminine-coded terms ($r = -0.33$), and show a weaker negative correlation with Masculine-coded terms ($r = -0.29$). In the Series line specifically, the correlation between Gender Scores and Feminine-coded terms is weak and positive ($r = +0.09$), while the correlation with Masculine-coded terms is moderate and positive ($r = +0.31$), suggesting a modest association between masculine language and more masculine-coded models.

Fourth, language-based coding also influences Gender Scores through color naming. Using a difference-in-effect-size analysis across gender categories (Hyper-Feminine, Feminine, Masculine, Hyper-Masculine), we found that Feminine and Hyper-Feminine color names were significantly associated with lower Gender Scores. Masculine-coded colorways, however, showed no consistent directional effect. This asymmetry suggests that names for feminine-coded colorways (e.g., 'Pink Sand' with a rough hex code of #d7c1c1 and 'Starlight' with a rough hex code of #e4e2de) had a stronger influence on reducing a model's Gender Score than names for masculine-coded colorways (e.g., 'Midnight' with a rough hex code of #44464c, *'Graphite'* with a rough hex code of #45433f)) did in increasing it. Notably, the effect size of the color names themselves often surpassed the effect size of the corresponding color hex codes on the Gender Score.

Fifth, an inverse correlation is observed between Innovation and Gender Scores. In the Series line, as Gender Scores become more Masculine, Innovation Scores trend downward ($r = -0.30$). This pattern becomes more extreme in the SE and Ultra lines, both showing a perfect negative correlation ($r = -1.0$), though these values are based on just two data points each. While not statistically generalizable, these results indicate that highly Masculine-coded models often receive lower Innovation Scores. When using Alternate B—which includes all app types, including watch faces—the correlation in SE and Ultra flips from a perfect negative ($r = -1.0$) to a perfect positive ($r = +1.0$). This suggests that when more elements are included, particularly lock screen features which display some version of time keeping, the overall Masculinity rating increases. In contrast, Alternate A—which includes only watch bodies and bands—produces correlation patterns nearly identical to the Primary Gender Score ($r = -0.145$).

However, this aggregate trend masks differences across product lines and score variants. In the Series line, for instance, Gender and Innovation Scores are weakly or inversely correlated (e.g., $r = -0.15$ using Alternate A). In contrast, SE and Ultra models display a perfect negative correlation ($r = -1.0$) when using the Primary or Alternate A scores, but flip to a perfect positive correlation ($r = +1.0$) under Alternate B. These shifts suggest that the inclusion of app-based features—particularly watch faces—can significantly alter gender coding, and by extension, the apparent relationship between gender and innovation. While these correlations should be interpreted with caution due to limited data points in the SE and Ultra lines, they point to the sensitivity of gendered readings to different configurations of design elements.

5 Discussion

Our findings offer implications for Human-Computer Interaction and Design (HCI&D) theory by providing empirical evidence of the entangled nature of innovation and masculinity. While our analysis centers on a single interactive consumer product, the dataset spans a decade of development and captures thousands of design, engineering, and marketing decisions across three product lines (Series, SE, Ultra) within the Apple Watch product group. This longitudinal scope offers a rare empirical context for examining how gender and innovation are co-constructed over time in an interactive consumer product. However, we do not claim that our findings are universally generalizable to interactive consumer products as a category. Rather, given that our focal product was developed by one of the world's most profitable and influential technology companies—and one that is consistently praised for leading interactive design and innovation [70, 73, 78]—we argue that it serves as a uniquely generative case for surfacing structural patterns that merit further critical inquiry within the field of HCI&D. Said another way, as a site where commercial aspirations, technological decisions, and cultural values converge, engineering design broadly—and in this case, the Apple Watch specifically—offers a revealing lens into how dominant ideas about gender and innovation become materially and rhetorically embedded in everyday interactive consumer products.

In the subsections that follow, we offer interpretations on key patterns identified in our results that clarify how innovation and gender norms have been jointly articulated, differentiated, and stratified through the Apple Watch product group over time.

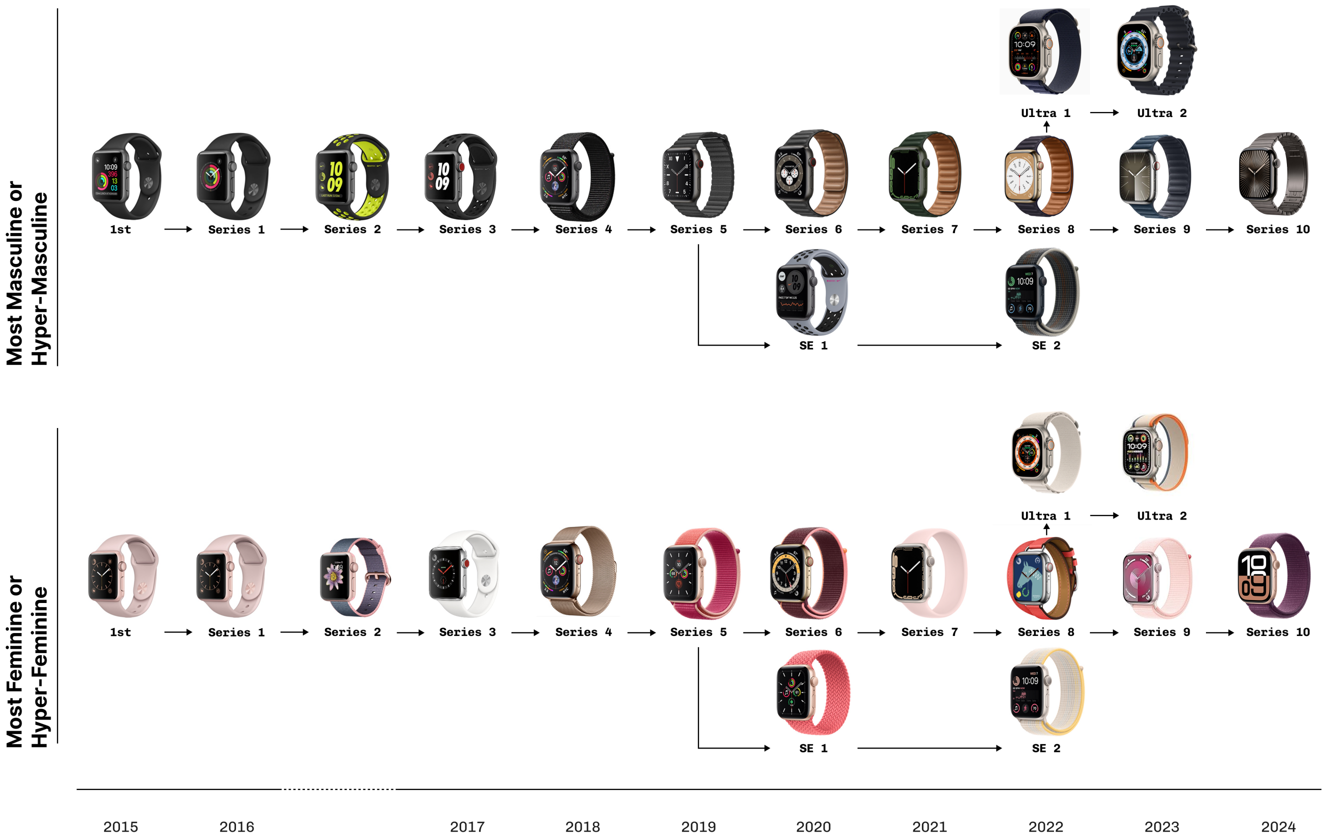
5.1 Product Line Stratification

The introduction of the Apple Watch SE (2020) and Ultra (2022) product lines reveal how Apple has seemingly stratified its product ecosystem in ways that align with normative gender hierarchies. Across these product lines, innovation is seemingly tiered—more technically advanced or robust models are positioned as more innovative, and tend to align with masculine-coded aesthetics and functions. As a result, the relationship between innovation and gender does not just unfold within product lines but is structured across them.

5.1.1 ***Series Line.*** The Series line, Apple's flagship product tier, demonstrates a modest inverse relationship between innovation and gender. Specifically, we observe a correlation of $r = -0.26$ between Innovation and Gender Scores, indicating that as models became less technically innovative, they tended to receive more Masculine-coded Gender Scores. While this correlation is not especially strong in isolation, it gains significance when viewed alongside the SE and Ultra lines, which each show a perfect negative correlation ($r = -1.0$). Taken together, these patterns suggest that masculinity is not just a stylistic overlay but may function as a structural response to stagnation in technical advancement, especially within the Series line, where year-over-year innovation seemingly became more difficult to sustain. Rather than directly co-occurring, gender and innovation appear to operate as counterweights in the Series line: as technical improvements slow or plateau, design elements coded as Masculine increase in prominence, perhaps to sustain perceptions of novelty, advancement, or value.

5.1.2 ***SE Line.*** The SE line occupies a more complex position—its focus on utility and affordability could be read as an attempt at gender neutrality, but as Bornstein [9] argues, such attempts at androgyny often imply masculinity or inadvertently reinforce rather than challenge gender binaries. Thus, forcing a binary-reading, the SE's omission of features like the Always-On Display and blood oxygen monitoring subtly reinforces a masculine utilitarian narrative, based on affordability and basic functionality. However, features such as the temperature sensor for menstrual tracking and the inclusion of health monitoring sensors (EKG/blood 02) challenge traditional masculine narratives by integrating feminine-coded applications. Complicating matters still, with the introduction of the Ultra line, the Series line seemingly became the mid-tier option, positions the SE seems to be the most 'neutral' option of the three, while still being a complicit in a system that reinforces a gendered hierarchy where more "premium" equals more masculine. Regardless of whether the SE is most masculine or feminine codes, its introduction signaled segmentation of features across a the product group which perpetuates a hierarchy. And how this hierarchy is constructed in the Ultra model especially, reveals a prioritization of Masculine- and Hyper-Masculine coded features (e.g., water resistance) in premium models and relegation or highlighting of Feminine and Hyper-Feminine coded features (e.g., menstrual tracking) to secondary tiers, like the SE.

5.1.3 ***Ultra Line.*** The Ultra line represents the apex of this hierarchy. Premium features consistently align with masculine-coded characteristics: larger sizes, "aerospace-grade" materials, and advanced technical capabilities. Nearly every attribute emphasized in Ultra landing pages—ruggedness, durability, enhanced battery life, precision tools, aerospace-grade materials—aligns with masculine-coded traits. This is further reflected in our data: no Ultra models achieved a Feminine or Hyper-Feminine Gender Score, despite the



Most Masculine and Feminine Designs by Release Year
All Apple Watch Models (2015–2024)

Figure 11: Highest- and lowest-scored industrial design variants across the Apple Watch Series, SE, and Ultra lines (2015–2024). Discretized by product line (from top to bottom: Ultra, Series, and SE) the top rows show the most Masculine- or Hyper-Masculine-coded variation released each year; the bottom rows show the most Feminine- or Hyper-Feminine-coded variation from that same year. Arrows indicate chronological release order and product group segmentation. When multiple watches received identical high or low scores, the figure includes a representative model to avoid duplication with previously shown examples. See Fig. 9 for each model's full Gender Score and ranges. Product Marketing Images: © Apple Inc..

inclusion of colorways or collaborations that might otherwise signal femininity. Instead, the Ultra line dramatically reinforces masculine norms through what could be termed 'innovation-as-ruggedness'—emphasizing features such as enhanced water resistance, larger screens, and extended battery life. These features, while technically innovative compared to the Series and SE lines, align closely with traditionally masculine values of endurance, power, and technical mastery.

5.2 Alternate Gender Score Metrics

To assess the robustness of our findings, we examined two alternate versions of the Gender Score (Alternates A and B; see Fig. **??**). These variants were designed to test how including or excluding different product features would affect the relationship between gender and innovation. Specifically, Alternate A includes only industrial design elements (watch bodies and bands), while Alternate B includes bodies, bands, and applications, but excludes lock screens (i.e., watch faces). Across all three scoring variants—including the Gender Score, which incorporates bodies, bands, and *all* applications—we found that correlations with Innovation Scores were relatively weak and varied by product line. In the Series line, the correlation between Gender and Innovation Scores ranged from $r = +0.04$ to $+0.10$, indicating a very weak positive relationship regardless of scoring method. In the SE line, the relationship was more consistently negative, with correlations ranging from $r = -0.18$ (Primary Gender Score) to $r = -0.34$ (Alternate B). In the Ultra line, correlations were positive and moderately strong, ranging from $r = +0.63$ (Primary) to $r = +0.72$ (Alternate A).

These results complicate any singular claim about the relationship between innovation and masculinity. While the Ultra line suggests that more Masculine-coded models may also be perceived as more innovative, the SE line shows the opposite trend. Meanwhile, the Series line reveals little relationship between the two.

Overall, this variation across scoring methods and product lines suggests that while certain patterns persist, especially within a given line, the relationship between gender and innovation is not fixed. Rather, it depends on which aspects of the product are emphasized and how masculinity is materially and rhetorically encoded in design.

5.3 Absence of Femininity

Perhaps the most striking pattern in our data is the near-total absence of models achieving Feminine or Hyper-Feminine Gender Scores. This is true even when feminine-coded elements are visibly present—most notably in colorways, materials, or special collaborations such as the long-term Hermès collaboration. These elements, while suggestive of femininity, were insufficient to produce an overall feminine-coded configuration once the body, band, and application were evaluated as a whole.

This absence is not simply a quirk of our scoring system; rather, it points to a deeper structural issue. It suggests that the frameworks we use to measure innovation in consumer technology are not gender-neutral. Instead, they appear to reward and amplify design features that align with masculine norms—such as: ruggedness, technical sophistication, and scale—while rendering feminine-coded qualities insufficient to register as innovative within the same system. In other words, when innovation is equated with functionality, endurance, or performance, and these traits are themselves culturally gendered as masculine, the metrics used to assess innovation will inherently skew toward masculine-coded designs. As a result, feminine-coded contributions, even when they exist, struggle to be legible as innovation within this evaluative structure.

This aligns with longstanding critiques in feminist STS and design theory, which have shown how dominant understandings of technology tend to reflect and reproduce patriarchal values under the guise of objectivity or neutrality. Scholars such as Wajcman [79, 80], Haraway [33], Cockburn [17], Oudshoorn and Pinch [48], Faulkner [29], and more have extensively documented and studied the how gender manifests in the design, development, and evaluation of technology. Our findings echo and extend this critique by demonstrating how gender norms are embedded not just in form or function, but in the very metrics by which technological progress is defined.

5.4 Public-Facing Communications

The marketing language and visual presentation further reinforce gendered patterns. Over time, masculine-coded terminology increased while feminine-coded language nearly disappeared by Series 9 (dropping from 0.013% to 0.00%). Even attempts at semantic diversity through color naming schema did not overcome the dominant masculine coding embedded in the product design. Said another way, this entanglement of innovation and gender manifests particularly clearly in the product marketing. Our analysis revealed that naming conventions had a significant effect on gender coding, with feminine and hyper-feminine color names showing a stronger influence (-0.2) on Gender Scores compared to masculine-coded names. Yet these feminine signifiers remained superficial—appearing primarily in color names like 'Starlight' and 'Gold'—while the underlying product architecture continued to privilege masculine-coded features. This dynamic reflects what Van Oost has described as gender "inscriptions" [76] of gender in technology: while some aspects of a product may signal femininity through ornamentation or aesthetics, the underlying functionality and architectural logic of the product often remains rooted in masculine norms. What emerges is a form of dual encoding where design and marketing gestures toward gender inclusivity while the product architecture and design structurally preserves existing normative gender hierarchies.

5.5 Theoretical Implications

When we observe how the trendlines in Gender and Innovation Scores maintain opposing trajectories even as new product tiers are introduced, we understand how gender is embedded not only in product options but also in the very framework through which we conceptualize and strategically position technological advancement.

Like other dominant modi of operating, as Cindy Katz notes in her work on Minor Theory, these design decisions often operate below the level of conscious awareness [36]. Instead, Katz and other critical theorists argue that the dominant paradigm is inadvertent and unexamined and thus perpetuated [36]. While we cannot know if designers were explicitly aware of this gendering in the course of design engineering this product, evidence that the design of the product was gendered appears over time and across multiple dimensions including the correlation patterns, the titration of features across product group, and the systematic skew toward masculine-coded design elements.

These findings challenge HCI&D researchers and practitioners to reconsider how we conceptualize and measure innovation itself. The correlation patterns we observed aren't merely describing relationships between variables—they reveal how gendered assumptions inform the segmentation of product lines, the distribution of features, and the very metrics we use to assess technological 'progress.' When premium features consistently align with masculine-coded characteristics, and when gender 'neutrality' consistently defaults to masculine norms, we must question whether current approaches to product innovation can achieve genuine inclusivity without fundamentally restructuring how advancement is conceptualized and measured in product design.

Historically and socially co-constructed gender norms shape not only who designs technology, but also whose needs, values, and preferences are prioritized in innovation processes [35]. When we contextualize technology within feminist theory, we expand the definition of 'technology' to represent a form of social knowledge, power, cultural practices, and products [29, 80]. Technological development can be seen then as a result dependent on the distribution of power and resources between different groups in society [18]. The framing of innovation as a pursuit of dominance and technical achievement reinforces masculinity as the default perspective, marginalizing alternative approaches to design and innovation. Ultimately, a feminist analysis revealed how pervasive the gender binary is in the shaping of technology, which sociologist Wendy Faulkner argued is a "particularly visible instantiation of the still durable cultural equation between masculinity and technology" [29].

5.6 Limitations

As with any novel mixed-methods approach, we acknowledge the limitations of our study, particularly from a methodological standpoint.

5.6.1 ***Binary Coding.*** One core limitation of this study lies in our decision to adopt a binary coding framework for both gender and innovation.

Our four-point scale—spanning Hyper-feminine (-2), Feminine (-1), Masculine (+1), and Hyper-Masculine (+1)—was explicitly designed to evaluate the presence of *normative* gender presentations. This approach enabled us to trace how dominant gender scripts are encoded in product design and communication, particularly as they relate to external documentation, industrial design, technical specifications. However, we acknowledge that this schema does not account for non-binary, fluid, or otherwise gender non-conforming expressions. This was not an oversight but a conscious methodological decision aligned with our aim: to surface how mainstream commercial products such as the Apple Watch reinforce dominant gender binaries through their industrial design and rhetorical positioning. That said, we recognize this decision necessarily limits the interpretive space of our analysis and risks reproducing the very binary logic we seek to critique. We also acknowledge that in attending primarily to gender as it is culturally legible within hegemonic Western contexts, our analysis may obscure other situated or intersectional readings of gendered design. Our coding approach was shaped by scholarship on gender scripts in technology, including Seaborn's work on "hypermasculinity" and "normative masculinity" in HCI&D [61], as well as longstanding feminist critiques of design's alignment with masculine-normed values [10, 30, 79]. We see our framework not as a comprehensive account of gender in interactive product design, but rather, as a focused lens through which to assess whether and how normative gender performances are embedded in interactive consumer products.

A parallel limitation exists in our binary approach to coding innovation. Here, we assigned +1 to features that signaled normative technological advancement (e.g., increased battery life, new sensors), and −1 to those that remained unchanged, were removed, or regressed. This approach privileges *internal progression* within the product line and does not account for *external contextualization*: a feature deemed innovative within one release may have existed in competitor products or adjacent industries for years, for example. As such, our measure of innovation captures the internal logic of progress as framed by the Apple Watch product group, rather than a broader comparative or industry-wide standard. This allowed us to evaluate Apple's own narrative of innovation, but we recognize that it does not reflect a universal or exhaustive assessment of innovation as a construct.

Future work might expand both dimensions of this framework, gender *and* innovation, through more relational or comparative approaches. In particular, triangulating conceptual assignments around gender and innovativeness through competitor analysis, or (intersectional) user perspectives could offer deeper insight into how gender and innovation are co-produced in ways that move beyond a binary logic.

5.6.2 ***Word Inventories and Lexicon-Based Analysis.*** Another limitation in this study lies in the respective word inventories for innovation and gender-related terms. While collecting inventories of words that evoke gender has a decently long history with foundational studies such as the Bem Sex Role Inventory (BSRI) [8] and Personal Attributes Questionnaire (PAQ) [64], innovation has not been interrogated in the same way, rhetorically. So while, the BSRI and PAQ and their successors have long codified *communal*[20], *expressive* [58], or *warm* [47] words (e.g. "emotional" or "gentle") with women or femininity, and agentic [20, 47] and *instrumental* [58] words (e.g., "aggressive" or "rational") with men or masculinity, no work of the same kind was identified by the authors for stereotypes of innovation.

Additionally, researchers in the HCI&D and NLP communities have pointed out the limitations of relying on fixed lexicons and have advocated for using supervised learning techniques on datasets to identify relevant words outside of predefined inventories [26]. This approach is in reaction to two known challenges with fixed inventories. First, target analysis data often include terms that are not present in established inventories [20]. Second, stereotypes and associations evolve over time, making fixed lexicons less adaptable to these shifts [26].

Despite these limitations, we intentionally built on the strengths of established word inventories while extending them via NLP to construct a broader and more context-sensitive lexicons. Still, this remains a quantitative approach—one that, by design, forgoes the contextual nuance and interpretive flexibility that qualitative methods afford. To support future qualitative analyses and the types of analytic power they afford, we provide full OCR-processed PDFs of all landing pages used in our study in the supplementary materials.

5.6.3 ***Wayback Machine and Wikipedia.*** The Wayback Machine served a critical role in our study by capturing snapshots of landing pages at or near the time of each product release. These archives allowed us to examine how Apple framed each release at the moment of its introduction, helping us capture the brand's intended rhetorical and visual positioning at each release. To ensure consistency, we selected the first available Wayback Machine snapshot taken after the official release date in California, USA. However, we later realized a temporal oversight in our use of this tool. Namely, the Wayback Machine timestamps are recorded in Coordinated Universal Time (UTC), not Pacific Standard Time (PST), and we did not originally adjust for this time zone difference. As a result, in a few cases, the archived snapshot may have slightly preceded the intended local release time.

Beyond the *External Documentation* and *Industrial Design Specification Datasets* creation, the Wayback Machine also supported our reconstruction of technical specifications in order to create the *Technical Specifications Dataset*. Because Apple does not maintain a public, longitudinal archive of historical specifications, we relied on alternative sources such as Wikipedia. Wikipedia provided one of the most comprehensive records of Apple Watch specifications, with entries often citing archived Apple pages which we attempted to access via the Wayback Machine. Together, these sources allowed us to triangulate data and reconstruct a detailed timeline of product features across releases. We further cross-referenced our reconstructions with Sun et al. [69], which examines technical dimensions closely aligned with our study, including display, sensors, processor, battery, and connectivity. While this paper offered additional validation of attribute categories, it is not longitudinal and does not cover all releases in our scope.

Still, this triangulated approach comes with limitations. Many of the technical details we relied upon were derived from secondary sources, which may introduce inaccuracies or inconsistencies. Archived specification pages were not always available through the Wayback Machine, and Wikipedia's categorization of hardware and software features may reflect editorial interpretations rather than original manufacturer intent. As such, we acknowledge that our classification of technical attributes may contain small inaccuracies or rely on imperfect groupings that do not fully represent the product's engineering complexity.

5.6.4 ***Collected and Uncoded Data.*** A final limitation of this study concerns the broader scope of data we collected compared to the subset ultimately incorporated into our scoring models. Several attributes were gathered to ensure completeness but were either excluded from our Gender and Innovation Score calculations or used only in supplementary analyses (e.g., hex color codes). Others, such as application category and action, were not utilized due to constraints in analytic capacity.

For instance, we collected hex color codes from Bandbreite for the background of each watch image [5]. However, background color proved unreliable as a proxy for band color, especially in instances where bands included multiple hues or were inconsistently lit. Similarly, we extracted hex values for watch bodies using eyedropper tools on product images, but these were difficult to apply meaningfully in our scoring due to visual artifacts like lighting gradients and shadows. Ultimately, we relied on the color names visible in product descriptions, which served as sufficient referents for our gender coding. It is important to note that while these color names were used to represent hue, the Gender Scores assigned to them reflect the chromatic value itself, rather than inferring a rhetorical gender coding.

Likewise, while we manually logged each application's category and action based on landing page presentation, we ultimately opted to use only the application name in our Gender Score analysis.

Although these additional variables were not incorporated into our core scoring methods, some informed secondary analyses—such as our examination of how semantic naming conventions influenced perceptions of color (see Section 3). We have chosen to make the full dataset, including these unused or partially used attributes, publicly available, as we believe many of these incomplete or unused variables may offer fruitful avenues for future research.

6 Conclusion & Future Work

In this work, we sought to add to the growing body of literature on design's role in innovation by implementing the Gender Value System (GVS) to qualitatively code and quantitatively score the Apple Watch's industrial design, technical specifications, and product language across ten years of releases. Our goal was to examine the extent to which innovation is co-constructed with gender and to carefully interrogate how normative values become embedded in interactive consumer technologies over time. Our findings support the entanglement of gender and technological design, revealing how cultural constructions of masculinity and femininity shape both the presentation and reception of innovation.

In answering our research questions, we identify three main contributions:

First, we present a novel, longitudinal methodology for measuring how gender and innovation are co-constructed in product design, combining qualitative coding and quantitative scoring across product tiers, materials, features, and marketing language. This mixed-methods framework reveals normative design patterns that might otherwise remain implicit, and serves as a tool for analyzing how social norms like gender shape interactive technologies over time.

Second, we find that gender and innovation are unevenly distributed across the Apple Watch product group—not only in the technical features included, but also in how these features are framed, prioritized, and allocated across product lines. The Ultra line, for instance, consistently scores as both the most innovative and the most masculine-coded, while the SE line contains fewer innovations and maintains the least masculine presentation. This stratification embeds gendered value hierarchies directly into the architecture of the product group, assigning cultural and technical significance to particular models through a gendered lens.

Third, we show that language plays a disproportionate role in the gendering of consumer technology. Feminine-coded color names, in particular, exert a greater downward influence on Gender Scores than the corresponding color values themselves. This effect highlights how rhetoric and labeling practices—not just technical or visual features—can reinforce gendered assumptions about products.

Importantly, we recognize the limitations of a binary coding schema for gender. While our analysis focuses on dominant norms of masculinity and femininity, it also reifies the binary categories it seeks to critique. We see this work as a starting point for broader empirical inquiry—one that can be extended across product categories and markets, and invite others to build on our framework using qualitative, non-binary, intersectional, or relational approaches to evaluating gender in design.

Looking ahead, our future work aims to support "critical consciousness raising" [18, 39, 83] within the Human Computer Interaction and Design (HCI&D) community by developing tools and frameworks that help surface how norms like gender are embedded in design decisions, product hierarchies, and even the very metrics we use to define innovation.

Rather than offering definitive answers, our study contributes a feminist informed framework for interrogating the assumptions embedded in interactive product ecosystems—inviting researchers, designers, and educators to rethink what counts as innovation, and for whom. In that spirit, we conclude with questions to guide future inquiry:

- *What are ways to subvert and amplify gender inclusivity through the design of interactive consumer products?*

- *How can design metrics be reimagined to recognize feminine-coded or non-dominant contributions as innovative?*
- *What interventions might disrupt the alignment between masculinity, technical advancement, and product 'premium-ness'?*
- *How does product group stratification reproduce cultural narratives around gender and value?*
- *What does it mean to measure innovation through a post-gender binary lens?*
- *And, how might these dynamics differ across product categories, markets, or sociotechnical contexts?*

Acknowledgments

This work is made possible with support of the National Science Foundation, under Grant Nos. 2034448 and DGE-2125913. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Lastly, this work was largely supported by our undergraduate research assistants, Duy Vu, Sheila Mercado, Sakchhi Thapa and the Co-Design Lab's Fall 2024 Data Science Discovery Project cohort. Thank you for your invaluable contributions to the data collection and analysis. We also want to thank Coye Cheshire and the 2024 and 2025 cohorts of his Doctoral Research and Theory Workshop.

References

[1] William J. Abernathy and James M. Utterback. 1978. Patterns of Industrial Innovation. *Technology Review* 80, 7 (1978), 40–47.

[2] Joan Acker. 1990. Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations. *Gender & Society* 4, 2 (1990), 139–158.

[3] Allan Afuah. 2003. *Innovation management-strategies, implementation, and profits.* Oxford University Press, New York.

[4] Joan Palmiter Bajorek. 2019. Voice recognition still has significant race and gender biases. *Harvard Business Review* 10 (2019), 1–4.

[5] Bandbreite. 2025. Bandbreite: A Visual Guide to Apple Watch Bands. https://bandbreite.watch/ Accessed: 2025-04-20.

[6] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1301–1310. doi:10.1145/1753326.1753521

[7] Peter Bell. 2012. Content analysis of visual images. *SAGE visual methods: Interpretation and classification* 3 (2012), 31–57.

[8] Sandra L. Bem. 1981. Gender schema theory: A cognitive account of sex typing. *Psychological Review* 88, 4 (1981), 354–364.

[9] Kate Bornstein. 1994. *Gender outlaw : on men, women, and the rest of us.* Routledge, New York.

[10] Samantha Breslin and Bimlesh Wadhwa. 2015. Towards a Gender HCI Curriculum. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 1091–1096. doi:10.1145/2702613.2732923

[11] Mark Buchanan. 2024. Techno-optimism needs a reality check. *Nature Physics* 20, 2 (2024), 176–176.

[12] Savatore Bucolo and Judy Matthews. 2011. Design-led innovation-Exploring the synthesis of needs, technologies and business models. In *2011 Participatory Innovation Conference Proceedings*. University of Southern Denmark, Søderborg, Denmark, 351–354.

[13] Judith Butler. 1988. Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal* 40, 4 (1988), 519–531.

[14] Carolina Herrera. n.d.. Good Girl Blush Eau de Parfum. https://www.ulta.com/p/good-girl-blush-eau-de-parfum-pimprod2037334?sku=2607898. https://www.ulta.com/p/good-girl-blush-eau-de-parfum-pimprod2037334?sku=2607898 Retrieved from Ulta Beauty.

[15] Shyam R Chidamber and Henry B Kon. 1994. A research retrospective of innovation inception and success: the technology–push, demand–pull question. *International Journal of Technology Management* 9, 1 (1994), 94–112.

[16] John Chisholm. 2016. What is design-driven innovation? http://www.designforeurope.eu/what-design-driven-innovation Accessed: 2025-01-18.

[17] Cynthia Cockburn. 1988. *Machinery of Dominance: Women, Men, and Technical Know-How.* Pluto Press, Dover, NH.

[18] Sasha Costanza-Chock. 2018. Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. In *Design as a catalyst for change - DRS International Conference 2018,*, Vol. 2. Design Research Society, Limerick, Ireland, 529–540. doi:10.21606/drs.2018.679

[19] Caroline Criado Perez. 2019. *Invisible Women: Data Bias in a World Designed for Men.* Abrams Press, New York.

[20] Jenna Cryan, Shiliang Tang, Xinyi Zhang, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao. 2020. Detecting Gender Stereotypes: Lexicon vs. Supervised Learning Methods. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3313831.3376488

[21] John Danaher. 2022. Techno-optimism: An analysis, an evaluation and a modest defence. *Philosophy & Technology* 35, 2 (2022), 54.

[22] Simone De Beauvoir. 2014. The second sex. In *Classic and Contemporary Readings in Sociology*. Routledge, Oxon, 118–123.

[23] Dolce & Gabbana. n.d.. Dolce Garden Eau de Parfum. https://www.sephora.com/product/dolce-garden-P429971. https://www.sephora.com/product/dolce-garden-P429971 Retrieved from Sephora.

[24] Philippe Doneys, Kyoko Kusakabe, Evelyn F. Wamboye, Rebecca Elmhirst, Arul Chib, and Joyee Shairee Chatterjee. 2022. Gender, technology and development: reflections on the past, and provocations for the future. *Gender, Technology and Development* 26, 3 (2022), 285–294. doi:10.1080/09718524.2022.2153459

[25] Karin Ehrnberger, Minna Räs"anen, and Sara Ilstedt. 2012. Visualising Gender Norms in Design: Meet the *Mega Hurricane Mixer* and the *Drill Dolphia*. *International Journal of Design* 6, 3 (2012), 95–98.

[26] Jacob Eisenstein. 2017. Unsupervised learning for lexicon-based classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[27] Arturo Escobar. 2011. *Encountering development: The making and unmaking of the Third World.* Princeton University Press.

[28] Jan Fagerberg. 2006. Innovation: A Guide to the Literature. In *The Oxford Handbook of Innovation*. Oxford University Press, Oxford. doi:10.1093/oxfordhb/9780199286805.003.0001

[29] Wendy Faulkner. 2000. Dualisms, Hierarchies and Gender in Engineering. *Social Studies of Science* 30, 5 (2000), 759–792. doi:10.1177/030631200030005005

[30] Wendy Faulkner. 2000. The Power and the Pleasure? A Research Agenda for 'Making Gender Stick' to Engineers. *Science, Technology, & Human Values* 25, 1 (2000), 87–119.

[31] Benoît Godin and Joseph P Lane. 2013. Pushes and pulls: Hi (S) tory of the demand pull model of innovation. *Science, Technology, & Human Values* 38, 5 (2013), 621–654.

[32] Lucy Griffin, Katie Clyde, Richard Byng, and Susan Bewley. 2021. Sex, gender and gender identity: a re-evaluation of the evidence. *BJPsych Bulletin* 45, 5 (2021), 291–299.

[33] Donna Haraway. 1991. A Cyborg Manifesto. In *Simians, Cyborgs, and Women: The Reinvention of Nature.* Routledge, New York, 149–182.

[34] Yvonne Hirdman. 2003. In *Genus : om det stabilas föränderliga former.* https://api.semanticscholar.org/CorpusID:161439553

[35] Samantha Kang and Andy Dong. 2023. Gender-Based Social Revolutions and Their Effect on Technology Evolution: A Case Study of the Sewing Machine. In *ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* Vol. 6: 35th International Conference on Design Theory and Methodology (DTM). ASME, New York, V006T06A002. doi:10.1115/DETC2023-114613

[36] Cindi Katz. 1996. Towards minor theory. *Environment and Planning D: Society and Space* 14, 4 (1996), 487–499.

[37] Euiyoung Kim, Jaewoo Chung, Sara Beckman, and Alice M. Agogino. 2016. Design Roadmapping: A Framework and Case Study on Planning Development of High-Tech Products in Silicon Valley. *Journal of Mechanical Design* 138, 10 (2016), 101106.

[38] Eileen Leonard. 2003. *Women, technology, and the myth of progress.* Prentice Hall, Upper Saddle River, NJ.

[39] Jon A. Leydens and Juan C. Lucena. 2018. Transforming Engineering Education and Practice. In *Engineering Justice: Transforming Engineering Education and Practice.* John Wiley & Sons, Inc., Hoboken, NJ, 197–241.

[40] Caitlin Lustig, Maya A Kaneko, Meghna Gupta, Kavita Dattani, Audrey Desjardins, and Daniela Rosner. 2024. Porous by Design: How Childcare Platforms Impact Worker Personhood, Safety, and Connection. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1336–1349. doi:10.1145/3643834.3661552

[41] Dora Marinova and John Phillimore. 2003. Models of innovation. *The international handbook on innovation* 1 (2003), 2164.

[42] Dana McKay, Huiwen Zhang, and George Buchanan. 2022. Who am I, and who are you, and who are we? A Scientometric Analysis of Gender and Geography in HCI. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery,

New York, NY, USA, Article 396, 19 pages. doi:10.1145/3491102.3502106
[43] Gloria Moss. 2017. *Gender, design and marketing: How gender drives our perception of design and marketing.* Routledge, London.
[44] William M. Newman, Alex S. Taylor, Christopher R. Dance, and Stuart A. Taylor. 2000. Performance targets, models and innovation in interactive system design. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York City, New York, USA) *(DIS '00)*. Association for Computing Machinery, New York, NY, USA, 381–387. doi:10.1145/347642.347796
[45] Anna Offenwanger, Alan John Milligan, Minsuk Chang, Julia Bullard, and Dongwook Yoon. 2021. Diagnosing Bias in the Gender Representation of HCI Research Participants: How it Happens and Where We Are. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 399, 18 pages. doi:10.1145/3411764.3445383
[46] Ruth Oldenziel. 1999. *Making technology masculine: men, women and modern machines in America, 1870-1945.* Amsterdam University Press.
[47] Jahna Otterbacher, Jo Bates, and Paul Clough. 2017. Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6620–6631. doi:10.1145/3025453.3025727
[48] Nelly Oudshoorn and Trevor Pinch. 2003. *How Users Matter: The Co-Construction of Users and Technology.* MIT Press, Cambridge, MA.
[49] Rik Raes, Saskia Lensink, and Mykola Pechenizkiy. 2024. Everyone deserves their voice to be heard: Analyzing Predictive Gender Bias in ASR Models Applied to Dutch Speech Data. *arXiv preprint arXiv:2411.09431* (2024).
[50] Mihail C Roco and William Sims Bainbridge. 2002. Converging technologies for improving human performance: Integrating from the nanoscale. *Journal of nanoparticle research* 4 (2002), 281–295.
[51] Jennifer A Rode, Yifan Feng, Annapoorni Chandrashekar, Sonia Andreou, and Andri Ioannou. 2024. Spools and Sparks: The Role of Materiality in Computational Making with E-textiles and BBC Micro: bit. In *Proceedings of the Halfway to the Future Symposium.* 1–11.
[52] Viktor & Rolf. 2024. Spicebomb Eau de Toilette. https://us.viktor-rolf.com/fragrance/spicebomb-eau-de-toilette-VKR_030.html Accessed: 2025-01-16.
[53] Gillian Rose. 2022. Visual methodologies: An introduction to researching with visual materials. (2022).
[54] Alyssa Ryan, Francis Tainter, Cole Fitzpatrick, Jennifer Gazzillo, Robin Riessman, and Michael Knodler. 2022. The impact of sex on motor vehicle crash injury outcomes. *Journal of Transportation Safety & Security* 14, 5 (2022), 818–842.
[55] Anup Sathya and Ken Nakagaki. 2024. Attention Receipts: Utilizing the Materiality of Receipts to Improve Screen-time Reflection on YouTube. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–16.
[56] Matthew N. Saunders, Carolyn C. Seepersad, and Katja Hölttä-Otto. 2011. The Characteristics of Innovative, Mechanical Products. *Journal of Mechanical Design* 133 (2011), 021009. doi:10.1115/1.4003409
[57] Jan Cornelius Schmidt. 2007. Normativity and Innovation: An Approach to Concepts of Innovation from the Perspective of Philosophy of Technology. In *2007 Atlanta Conference on Science, Technology and Innovation Policy.* IEEE, Atlanta, GA, 1–8. doi:10.1109/ACSTIP.2007.4472880
[58] Stephen A Schullo and Burton L Alperson. 1984. Interpersonal phenomenology as a function of sexual orientation, sex, sentiment, and trait categories in long-term dyadic relationships. *Journal of Personality and Social Psychology* 47, 5 (1984), 983.
[59] Joseph A. Schumpeter. 1947. *Capitalism, socialism and democracy.* Harper & Brother, London.
[60] Joseph A Schumpeter. 2021. *The theory of economic development.* Routledge, London.
[61] Katie Seaborn. 2023. Interacting with Masculinities: A Scoping Review. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 179, 12 pages. doi:10.1145/3544549.3585770
[62] Helen Sharp, Jennifer Preece, and Yvonne Rogers. 2019. *Interaction Design : Beyond Human-Computer Interaction.* John Wiley & Sons, Inc, Indianapolis, IN.
[63] John Spacey. 2019. 34 Words For Innovation. https://simplicable.com/productivity/words-for-innovation. Accessed: 11 March 2024.
[64] Janet T. Spence, Robert Helmreich, and Joy Stapp. 1975. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology* 32, 1 (1975), 29–39.
[65] Katta Spiel. 2021. "Why are they all obsessed with Gender?" — (Non)binary Navigations through Technological Infrastructures. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 478–494. doi:10.1145/3461778.3462033
[66] Denny L. Starks, Tawanna Dillahunt, and Oliver L. Haimson. 2019. Designing Technology to Support Safety for Transgender Women & Non-Binary People of Color. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion* (San Diego, CA, USA) *(DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 289–294. doi:10.1145/3301019.3323898
[67] Statista. 2024. *Worldwide smartwatch shipment market share from 2014 to 2024.* https://www.statista.com/statistics/910862/worldwide-smartwatch-shipment-market-share/
[68] Anneli Stenberg. 2016. What do people associate with innovation and why? https://www.diva-portal.org/smash/get/diva2:1057358/FULLTEXT01.pdf.
[69] Xianghao Sun. 2024. A comprehensive engineering design analysis of Apple Watch as a smart wearable device. *Applied and Computational Engineering* 71 (2024), 52–57.
[70] Bucic Tania and Gaganpreet Singh. 2018. Apple watch: managing innovation resistance. (2018).
[71] Margaret Taylor and Andrew Taylor. 2012. The technology life cycle: Conceptualization and managerial implications. *International journal of production economics* 140, 1 (2012), 541–553.
[72] The Independent. 2018. *Apple's new iPhone XS Max is too big for women's hands, say critics.* Independent Digital News and Media Limited. https://www.independent.co.uk/life-style/women/apple-iphone-xs-max-new-too-big-size-women-sexist-feminist-a8537171.html Accessed: 2025-01-20.
[73] Paul Trott. 2008. *Innovation management and new product development.* Pearson education.
[74] Valentino. n.d.. Valentino Uomo Intense Eau de Parfum. https://sensabeauty.com/products/valentino-uomo-intense-edp. https://sensabeauty.com/products/valentino-uomo-intense-edp Retrieved from Sensa Beauty.
[75] Ellen Van Oost. 2003. Materialized gender: How shavers configure the users' femininity and masculinity. *How users matter: The co-construction of users and technologies* (2003), 193–208.
[76] Miriam Van Tilburg, Theo Lieven, Andreas Herrmann, and Claudia Townsend. 2015. Beyond "pink it and shrink it": perceived product gender, aesthetics, and product evaluation. *Psychology & Marketing* 32, 4 (2015), 422–437.
[77] Roberto Verganti. 2008. Design, meanings, and radical innovation: A metamodel and a research agenda. *Journal of product innovation management* 25, 5 (2008), 436–456.
[78] Roberto Verganti. 2009. *Design driven innovation: changing the rules of competition by radically innovating what things mean.* Harvard Business Press.
[79] Judy Wajcman. 1991. *Feminism Confronts Technology.* Penn State Press, University Park, PA.
[80] Judy Wajcman. 2004. *TechnoFeminism.* Polity, Malden, MA.
[81] Zixuan Wang, Jiawen Huang, and Costa Fiammetta. 2021. Analysis of Gender Stereotypes for the Design of Service Robots: Case Study on the Chinese Catering Market. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1336–1344. doi:10.1145/3461778.3462087
[82] Candace West and Don H Zimmerman. 1987. Doing gender. *Gender & society* 1, 2 (1987), 125–151.
[83] Geling Xu, Aerin Benavides, Angela Calabrese Barton, Edna Tan, Selena Bliesener, Gina DiFrancesco, and Scott Calabrese Barton. 2022. Critical consciousness in engineering for sustainable communities: A justice-oriented approach connecting schools and communities. *Connected Science Learning* 4 (2022). Issue 1. https://www.nsta.org/connected-science-learning/connected-science-learning-january-february-2022/critical-consciousness

A Appendix

This codebook enabled authors to assess how gender is encoded in Apple Watch product design and text. Its application enabled researchers to identify and catalogue design elements and patterns that communicate gender through visual representation and text on Apple Watch landing pages. The unit of analysis is a visual depiction of an Apple Watch sourced from Apple's website.

Four gendered categories were established and a numeric value was associated with each code: Hyper-Masculine (2), Masculine (1), Feminine (-1), or Hyper-Feminine (-2). Codes were applied to nine attributes of the watch using an Attribute x Code table that included a discrete list of all variations of each attribute and a gender code assigned to each variation.

In the table below, the orienting language rows summarize the authors' theoretically-grounded ideas that guided the code classifications. The visual cues rows describe specific aspects of the image that we expected to see in each gender category given that category's orienting language. The examples rows provide excerpts from the Attribute x Code table used by the research assistants.

For additional details and to access all datasets used in this study, please visit: https://codesign.berkeley.edu/papers /ballestas-norms-innovation-gender.

Figure 12: Apple Watch Band Gender Codes and Attributes Table

Attribute	Hyper-Masculine (+2)	Masculine (+1)	Feminine (-1)	Hyper-Feminine (-2)
Band Material	Materials, style, and clasp designs that emphasize technical superiority and rugged capability; styling celebrates its own robustness, durable against adventure and risk	Materials, style, and clasp designs that prioritize reliable functionality with professional polish; technical orientation, sporty	Materials, style, and clasp designs that balance softness with functionality; conveys approachable sophistication through tactile comfort.	Materials, style, clasp designs that celebrate ornamentation and delicate complexity; conveys artisanal luxury through intricate detail.
Visual Cues	Heavy-duty clasps, tactical gear style, reinforced attachment points, magnetic/mechanical closures.	Classic buckles, professional metal links, clean sport materials, traditional leather, straight-forward closures.	Woven soft materials, slim elegant straps, double-wrap designs, refined buckles, gentle, organic, fluent curves.	Delicate chains and links, leather, decorative closures, intricate patterns, jewelry-inspired hardware.
Examples	Titanium Milanese Loop, Deployment Buckle, Alpine Loop	Classic Buckle, Sport Loop, Leather Loop	Gourmette Double Tour, Modern Buckle, Attelage Double Tour	(None)
Band Color	Colors unapologetically assertive, emphasizing raw power; masculinity as high-performance machinery,warning signs, and hazard stripes.	Colors effortlessly authoritative, conveying competence without showmanship; masculinity as polished mahogany, quiet confidence.	Softer and harmonious colors; blends well with other colors; femininity as a whisper rather than a shout.	Bright and bold colors, manifesting operatic femininity; hot pinks, exclamation points, and rose gold extremes.
Visual Cues	Very dark colors, high-vis greens, warning oranges.	Medium blues and grays, traditional reds, neutral browns, silver	Pastels, beiges, creams, pure white.	Bright pinks and purples, high red and pink values.
Examples	Black & Volt, Graphite, Midnight.	Camel, Cape Cod Blue, Clay.	California Poppy, Chalk Pink, Gold. & Bordeaux	Rose, Extrême & Rose Azalée, Electric Pink
Band Size	Sizes designed exclusively for male physiological characteristics, emphasizing large size	Sizes developed with male characteristics as the baseline; assumes male averages.	Sizes acknowledging and designing for female physiological characteristics, accommodating smaller wrists.	Sizes requiring female physiological characteristics to function; bracelet-like designs
Examples	46 mm (Big)	45 mm	41 mm	40 mm & 42 mm (Little)

Figure 13: Apple Watch Body Gender Codes and Attributes Table

Attribute	Hyper-Masculine (+2)	Masculine (+1)	Feminine (-1)	Hyper-Feminine (-2)
Body Material	Materials convey advanced technology and superiority, hardness, durability; seen as military or industrial grade	Materials convey practicality along with durability and a professional polish; uses laboratory grade or brushed stainless steel.	Materials convey comfort, softness, practicality, visually appealing, classy	Materials are visually eye-catching, conveys sense of high-end, luxurious feel
Visual Cues	Titanium, carbon fiber, reinforced composites or metals, military-grade alloys.	Aluminum, stainless steel, industrial polymers, brushed metals.	Gold, ceramic, porcelain, polished stone, pearl finishes.	Rose gold, crystal & diamond-like finishes, mirror-polished metals.
Examples	Titanium.	Stainless steel.	18 Karat yellow gold.	18 Karat rose gold.
Body Color	Body colors convey assertive masculinity; emphasizes strength, risk, power, high-performance machine quality	Body colors convey authority, masculine confidence, classiness	Body colors convey softness, style, warmth, confidence, blend well with others	Body colors are deliberately spectacular, bold, unapologetic, passionate femininity; hot pinks
Visual Cues	Very dark colors, high-vis greens, warning oranges accents	Medium blues and grays, traditional reds, neutral browns, silver.	Pastels, beiges, creams, pure white.	Bright pinks and purples, high red and pink values.
Examples	Space Black, Graphite.	Natural, Silver.	Pink, White.	Rose Gold.

Figure 14: Apple Watch Applications Gender Codes for App Attributes Table

Attribute	Hyper-Masculine (+2)	Masculine (+1)	Feminine (-1)	Hyper-Feminine (-2)
App Functionality	Apps with core functions or purposes related to performance, technical precision, and competition. Includes advanced metrics, technical tools, and competitive tracking. User experience is data-heavy, technical, and performance-oriented.	Apps with core functions or purposes related to utility, efficiency, and practical function. Includes basic tools, information delivery, and organization. User experience is straightforward, professional, and reliable.	Apps with core functions or purposes related to wellness, lifestyle balance, and social connection. Includes mindfulness apps, activity tracking, and social sharing. User experience is supportive, community-oriented, and offers gentle guidance.	Apps with core functions or purposes related to personal transformation and emotional well-being. Includes female-specific health, beauty/lifestyle tracking, and guided wellness. User experience is highly personalized and emotionally supportive.
Visual Cues	Heavy emphasis on technical features, precision-oriented layouts, data visualization, performance dashboards	Clean and minimal interfaces, professional color schemes, focus on usability and organization.	Soft colors, intuitive layouts, features designed for social or lifestyle balance.	Ornate and personalized interfaces, bright and attention-grabbing aesthetics, focus on emotional connection.
Examples	Dexcom G5, MEATER® SmartMeat Thermometer, advanced fitness tracking apps.	Calculator, basic productivity tools, news delivery apps.	Breathe, Pocket Yoga, activity ring trackers.	Cycle Tracking, guided meditation apps, beauty and wellness tracking tools

Figure 15: Gender and Innovation Word Inventories used for External Document Content Analysis. Words drawn from [8, 20, 47, 58, 64]. For the full list of words used in this analysis, please visit https://codesign.berkeley.edu/papers/ballestas-norms-innovation-gender

Feminine (F) Gendered Words	NLP Expanded Feminine (F) Gendered Words	Masculine (M) Gendered Words	NLP Expanded Masculine (M) Gendered Words	Innovation Words	NLP Expanded Innovation Words
Abbess	Adorn	Abbot	Advanced	Advancement	Adapt
Actress	Adorned	Able_seaman	Adventurous	Betterment	Adaptation
Adulteress	Adorning	Adonis	Aesthetic	Breakthrough	Adapted
Adventuress	Adornment	Aircraftsman	Aesthetics	Challenge	Adapting
Amazon	Aesthetic	Aircrewman	Athletic	Change	Adaptive
Ambassadress	Aesthetically	Alderman	Athletics	Creative	Advance
⋮	⋮	⋮	⋮	⋮	⋮
Women	Woven	Yachtsman	Versatility	Transformative	Visionary
Wonder_woman	YSL	Yardman	Victorino	Visionary	Visualizing