

Content QA in the era of AI

Andrew Sales

*Chief Content Architect
Bloomsbury Publishing Group plc*

NEC SWS Publishing Technology Conference
5th November 2024

About me

- Single-source digital publishing
- Content Architecture team lead
- ISO Schematron project editor & WG convenor
- BSI IST/41 chair (document description and processing languages)



About Bloomsbury

- Established 1986
- Two main divisions



- Our team works mainly with A&P, specifically digital (BDR)



About BDR

- Wide range of content types:
 - Text-based (monographs; MRWs; legal)
 - Performed works (plays, screenplays)
 - Digital archive (Churchill Archive; rare books)
 - Multimedia (images, video, audio)
 - People; Museums & Exhibitions; Buildings etc
- ... and of subject areas
- Others' content too



Caveats

- This is not a case study!
- Not the product of an internal research project



Definitions

- “Content” - **structured data** (XML)
- “Quality” - **sufficient & consistent** to meet internal & consumer* need
- “Assurance” - (preferably at least semi-) automated means to ensure quality criteria are met

**consumer* could equally be machine e.g. discovery service etc.



“Content QA” reformulated

Ensuring, as far as possible by automated means, that the XML used to publish our products and metadata about them is sufficient and consistent enough to meet both internal and consumer need



“Content” vs “data”

- Documents or data?
- **Data:** machine-generated (predictable structure)
 - Often *machine-to-machine* (e.g. finance)
 - Easier to check with existing technology (schemas)
- **Content:** Human-authored (variable structure)
 - Not so easy to check...



Sufficiency: quality

REQUIREMENT:

“If a cited work is referred to in the body of the text, it must be linked to that work’s entry in the bibliography.”

EXAMPLE:

“This thought-provoking book (Thatcher, 1979)...”



Sufficiency: assurance

SOLUTION:

Search text for author surname + date

BUT text-matching (non-AI) tools can produce false positives, e.g.

“That decade also saw the UK elect its first female prime minister (Thatcher, 1979)...”

CONCLUSION: *Absence* of markup is hard(est?) to check for



The view from Prague



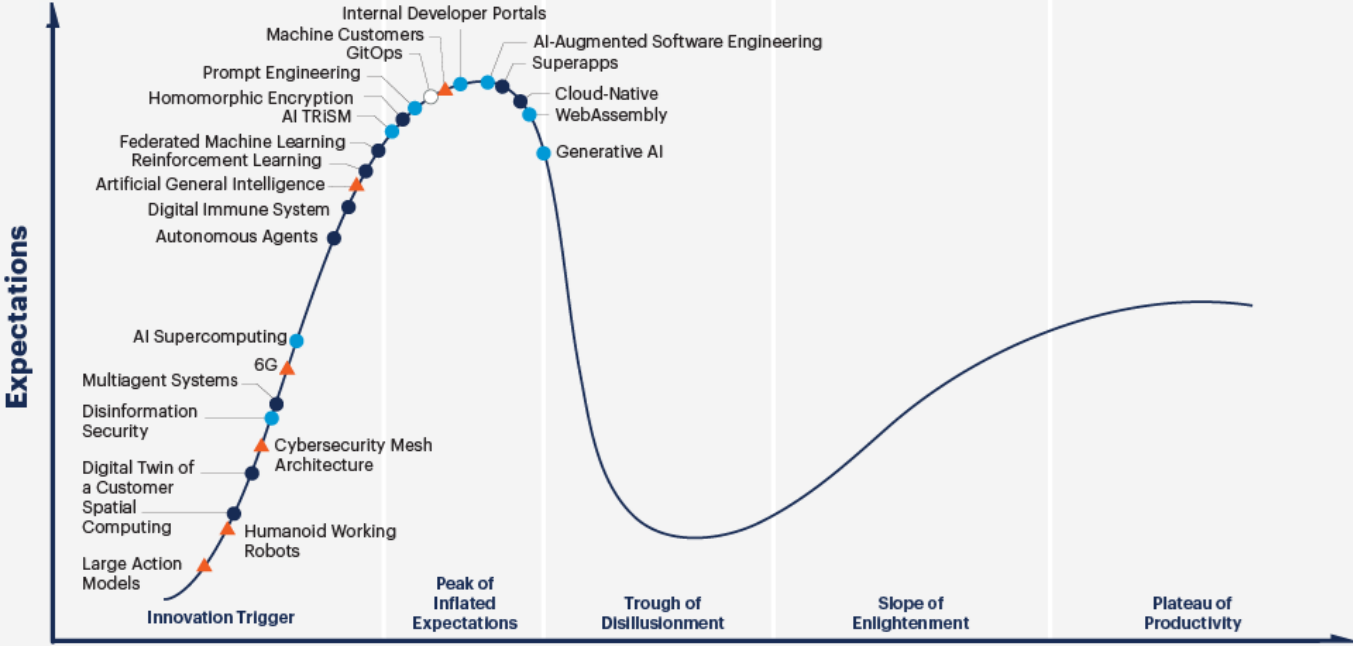
Credit: <https://www.flickr.com/photos/alwyn22>

The view from Prague

- [XML Prague](#) 2024
- [Marta Bartnicka \(Dolby\)](#)
- General consensus (cf. [Balisage](#)): better AI results from structured documents vs plain text
- Markup is designed to be machine-readable, after all
- Perils of hallucination!



Hype Cycle for Emerging Technologies, 2024



Plateau will be reached:

- less than 2 years
- 2 to 5 years
- 5 to 10 years
- ▲ more than 10 years
- ⊗ obsolete before plateau

Time

As of September 2024

Source: Gartner
 Commercial reuse requires approval from Gartner and must comply with the Gartner Content Compliance Policy on gartner.com.
 © 2024 Gartner, Inc. and/or its affiliates. All rights reserved. 3205434



An experiment

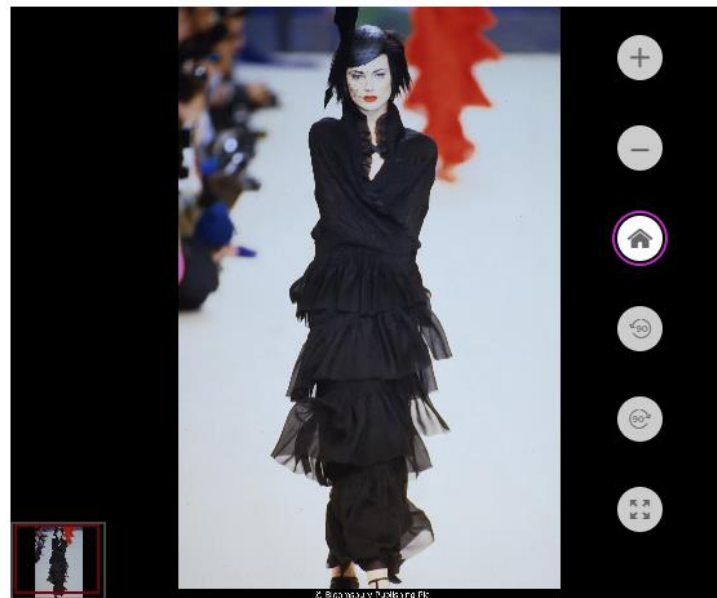
Alt text for images

A11y, AI and alt text

- Fashion Photography Archive
- 775K+ images
- AI cloud service for alt text
- Quality of results too low
- Use of AI itself entailed QA...

Chanel, Spring/Summer 1995

by Chanel (designer), Karl Lagerfeld (designer: person), Chanel (label), Niall McLnerney (photographer)



What QA tools can we use *now*?

<oxygen/> & Schematron ([Nadolu, Markup UK 2023](#))

- Schematron expresses business rules
- Content improvement API calls
- [SQF](#) to allow human oversight

```
<sch:rule context="shortdesc">
  <sch:assert test="ai:verify-content('Is active voice used?', .)"/>
  In the description we should use active voice.</sch:assert>
</sch:rule>
```

FIND

FIX

```
<sqf:fix id="rephrase">
  <sqf:description>
    <sqf:title>Reformulate the text to use active voice</sqf:title>
  </sqf:description>
  <sqf:replace match="text()" select="ai:transform-content('
    Reformulate to use active voice', .)"/>
</sqf:fix>
```



QA & AI

- Plausibility & reliability
- What is the quality metric and how is it enforced?
- Confidence in image alt text; hallucination
- “Who checks the checker?”



“Plateau of Productivity”?

- Much in place already
- Hinges on results AIs can provide
- Rapid improvement helps
- Ideal would be documents marked up by machine
 - The end of QA...?



Conclusion

- “All programmers are optimists” - Fred Brooks
- Automated subject indexing “just around the corner”
- Meantime: blend AI with human oversight
- QA productivity enhancer, not direct replacement



Thank you

Any questions?

