OWASP PDF Cookbook (v3.9.0)

1. Quick Start on Mac

Three steps to create a clone PDF of Version 1.1 of LLM AI Security & Governance Checklist from nothing:

1. Create a folder on your Mac desktop, `Cookbook`
2. Download `owasp_pdf` executable from this repository to `Cookbook` folder.
   (Note: in case Mac complains that it's not downloaded from App Store)
3. Run `owasp_pdf` with project `GOV` and language `en-US`
   (Note: period before /owasp_pdf)

```
$ cd ~/Cookbook
$ ./owasp_pdf -l GOV_en-US
    *** Processing GOV_en-US
    509 lines written to <your>/Desktop/Cookbook/llm-top-10-governance-doc/en-US/baseline/GOVAll_en-US.md
    29 page PDF created on <your>/Desktop/Cookbook/llm-top-10-governance-doc/en-US/baseline/GOVAll_en-US.pdf
    Processing time: 42 seconds
```

2. Project Code and Language Code

Project code is a random three-capital-letter word such as `GOV` or `LLM` assigned to every project in OWASP PDF project registration process (see 3. below). Project is typically associated with github repository. For example, `GOV` is associated with llm-top-10-governance-doc repository. As of this owasp_pdf version, registered project code and corresponding github repository are as follows:

OLM : olm
LLM : 1_1_vulns
GOV : llm-top-10-governance-doc

Language code consists of two parts ISO 639-2 language code and ISO 3166 region code joined by hyphen e.g. `en-US`. In the OWASP PDF system, there can be two or more language codes associated with one language. Three factors drive language variations:

▷ Font glyphs
▷ Word spelling
▷ Paper size

For example, we have `en-US` and `en-GB` because of paper size and spelling.
We use not ISO 639-2 language code `ja` or `ko` alone but `ja-JP` and `ko-KR` for Japanese and Korean although they are the only Japanese and Korean language codes respectively. `en-ZZ` is reserved as International English which uses `en-US` glyphs and spelling and `A4` paper size. `fr-CA` and `fr-FR` share the same glyphs and spelling, but they use different paper sizes: A4 for `fr-FR` and Letter for `fr-CA`. `en-ZZ` is the source of every localization.

ISO References
1. ISO 639-2 language code
2. ISO 3166 region code

3. Project and Language Registration

Registered project means the following:
▷ Project code is assigned
▷ The project code is associated with github repository
▷ `en-ZZ` markdown files and template parts are created
▷ `en-ZZ` and `en-US` languages are registered, and
▷ They are all built into `owasp_pdf` executable

Registered language means the following:
▷ Localization is ready to start. First invocation of `./owasp_pdf -l <lang code>` command generates a pristine PDF, which is exactly the same as en-ZZ PDF except for the page size, e.g. A4 for `en-GB` and Letter for `en-US`.

4. Localize MarkDown Files

The pristine PDF is created from the markdown files pre-populated in `en-ZZ` language. After the first execution of `./owasp_pdf -l <lang code>` command, no markdown files exist on your `<lang-code>` directory. Copy one or more of the markdown files from `en-ZZ` to your `<lang-code>` directory. After some edits, run `./owasp_pdf -l <lang code>` -- your localized markdown file(s) will be added to the PDF. Missing markdown files are picked up from `en-ZZ`.

5. Customize PDF

In the first `./owasp_pdf -l <lang code>` run, `custom_data_<proj code>_<lang code>.json` file is created under `/baseline` directory. The JSON file defines all the customizable parameters such as document title, font size of various parts of PDF. See `APPENDIX. custom pdf json` section.

- ▷ Set boolean: `doc_cover`, `doc_legal_notice`, or `doc_toc` to `false` to remove cover page, legal disclaimer, or table of contents respectively.
- ▷ Set `doc_title` and `doc_subtitles` to localize the cover page.
- ▷ Set `md_file_range` to remove specific markdown files from PDF creation process.
- ▷ Use different values of `md_file_range` to switch multiple markdown file sets. Any numbers between 0 and 999,999 (inclusive) can be set to `md_file_range`.

6. Technology Stack of PDF Creation

owasp_pdf is built on open source packages in Python. Implementation module of owasp_pdf of each technology layer is as follows:

parameter definition : custom_data_*.json (JSON)
interpretation : owasp_pdf (Python)
pdf generation : reportlab, fpdf2 (Python)
system language : Python

7. Minimalism for Simultaneous International Releases

Purpose of OWASP PDF system is to drive LLM Top 10 to the main stream of AI:

- ▷ Structural simplicity is a key to scalable publication and localization
- ▷ Consistent typographical quality across publications & localized PDFs demonstrates business integrity

What does minimalist design of OWASP PDF mean in the OWASP Top 10 documentation process? It means the following:

- ▷ As described above:
  - ▸ `owasp_pdf` contains everything you need to build PDFs from markdown files
  - ▸ User facing feature set is minimal aiming at consistent look&feel across publications and localization, but
  - ▸ New features can be added up to the point where the underlying open source packages (reportlab, fpdf2) support, which is pretty fancy PDF generation
- ▷ Document-level PDF structure is pre-defined: cover page --> toc --> legal notice/disclaimer, then your markdown files. toc and legal notice/disclaimer can be turned on/off in `custom_data json` file

8. Typographical Considerations

owasp_pdf typographical design rules are based on and quoted from:
UTTERICK'S PRACTICAL TYPOGRAPHY © 2010–24 Matthew Butterick

8.1 Bold or Italics

In-line text emphasis is not supported.

> Typographical rule: Bold and Italics
>
> Bold or italic—-think of them as mutually exclusive. That is the rule #1.
>
> Rule #2: use bold and italic as little as possible. They are tools for emphasis. But if everything is emphasized, then nothing is emphasized. Also, because bold and italic styles are designed to contrast with regular roman text, they're somewhat harder to read. Like all caps, bold and italic are fine for short bits of text, but not for long stretches.: Use bold and italic as little as possible.

8.2 Tables

Tables are not supported.

> Typographical rule: Tables
>
> For spreadsheet-style grids of numbers or other data. In the typewriter era, grids like this would've been made with tabs and tab stops. These days, you'd use a table.
>
> For layouts where text needs to be positioned side-by-side or floating at specific locations on the page. If making these is frustrating with the usual layout tools, try using a table.
>
> Cell borders are the lines around each cell in the table. Cell borders are helpful as guides when you're loading information into the table. They're less useful once the table is full. The text in the cells will create an implied grid. Cell borders can make the grid cluttered and difficult to read, especially in tables with many small cells.

8.3 Widow and Orphan Control

Widow and Orphan Control is not supported.

> Typographical rule: Widow and Orphan Control

Picture a paragraph that starts at the bottom of one page and continues at the top of the next. When only the last line of the paragraph appears at the top of the second page, that line is called a widow. When only the first line of the paragraph appears at the bottom of the first page, that line is called an orphan.

Widow and orphan control prevents both. Orphans are moved to the next page with the rest of the paragraph. To cure widows, lines are moved from the bottom of one page to the top of the next. It's a little more complicated than it sounds, because curing a widow cannot create a new orphan, nor vice versa.

8.4 First Line Indent

First-line indents are not used. Space is used between paragraphs.

Typographical rule: First Line Indent

First-line indents and space between paragraphs have the same relationship as belts and suspenders. You only need one to get the job done. Using both is a mistake.

8.5 Paragraph Tab and Ordered/Unordered Lists

Two space characters can be used at the beginning of a line to add one tab. Four spaces add two tabs; Six spaces add three and so on. Regular tabs are left tab stops.

Ordered lists begin with one or more consecutive digits followed by period character and one space character. Ordered lists are rendered as decimal tab stops. Regular paragraph tabs can be combined with ordered lists, e.g., Two spaces followed by digits, one period, and one space are rendered as one paragraph tab and an ordered list.

Unordered lists are rendered like the ordered lists. It begins with minus, plus, or asterisk character followed by one space character. Unordered list marker character is added at the beginning of the list. It's rendered as right tab stops at the marker character. Marker character is customizable - choice is "square", "circle", "diamond", and "triangle". Default is "square." Marker character cyclically changes as the paragraph tab depth increases, i.e., □ -> ■ -> ▫ -> ■ -> ▫ and so on.

Typographical rule: Bulleted and Numbered Lists

Are you still making bulleted and numbered lists by manually typing bullets or numbers at the beginning of each line?

In the 21st century, no one should be doing this task by hand. Manually formatted lists are a

waste of time and prone to error. Use automated lists.

NOTE: owasp_pdf does not support automated lists because the minimalist design overrides the rule of thumb.

Typographical rule: Tabs and Tab Stops

Tabs are used in bulleted and numbered lists to separate the bullet or number from the text. Tabs are also used in automatically generated tables of contents and tables of authorities to put the page numbers at the right edge of the table.

8.6 Justified Text

Default for body text and reference text is 'justified.' It's customizable usually 'left' if not ' justified.' Justified text is not supported for heavy ligature languages such as Hindi (hi-IN) and Bi-Di (ar-SA, he-IL).

Typographical rule: Justified Text

Justification is a matter of personal preference. It is not a signifier of professional typography. For instance, most major U.S. newspapers and magazines use a mix of justified and left-aligned text. Books, on the other hand, tend to be justified.

If you're using justified text, you must also turn on hyphenation to prevent gruesomely large spaces between words.

8.7 Font Style for Body Text

Serif is used for body text if the TrueType font is available. If not, Sans-Serif is used. Size and line pitch of body text is customizable.

Typographical rule: Font Style for Body Text

Please note: body text is the most common element of a document. Therefore, how the body text looks will have the most noticeable effect on the appearance of the document. Consequently, you should set up the body text first.

Start with font, point size, line spacing, and line length, because those four decisions will largely determine how the body text will look.

Though I'll stop short of calling it a rule, I strongly recommend using a serif font —not a sans

> serif font— for body text in print. Most books, newspapers, and magazines use serif fonts for body text. It's the traditional choice and still the best choice.
>
> On the web, body text can be in a sans serif or serif font. Sans serifs were once preferred for screen text because they rendered better on the lower-resolution screens of the past. (That's why most graphical user interfaces are built around sans serif fonts). But on today's screens, serif fonts look equally good.

8.8 Line Spacing/Pitch

Default line pitch for body text is 1.6 x font size for CJK fonts and 1.4 x otherwise. Line pitch is customizable.

> Typographical rule: Line Spacing
>
> For most text, the optimal line spacing is between 120% and 145% of the point size. Most word processors, as well as CSS, let you define line spacing as a multiple. Or you can do the math—multiply your point size by the percentage. (The text in this paragraph has line spacing of 110%. It's too tight.)

8.9 Line Length

Page margin is 1.0 inch each side. Line length is 'page width' minus two inches. Not customizable.

8.10 In-line Reference

Two or more reference links in one MarkDown text line are not supported. In-line reference links should be avoided. Recommended to re-write this:

> Consider both Split-View Data Poisoning and Frontrunning Poisoning attack vectors for illustrations.

to this:

```
> Consider both Split-View Data Poisoning (Ref.1) and Frontrunning Poisoning (Ref.2) attack vectors for illustrations.
>
> Reference Links
>   1. Split-View Data Poisoning
>   2. Frontrunning Poisoning
```

or this (if in-line reference is required):

```
    Consider these attack vectors for illustrations:
        Split-View Data Poisoning, and
        Frontrunning Poisoning
```

8.11 Text Color

Colors are used in two places in owasp_pdf system: link text and block6, light-blue and light-orange respectively. The colors are not customizable.

> Typographical rule: Color
>
> These days, color printers are ubiquitous and more writing is delivered on screen. So color has become a practical consideration.
>
> 1. On a page of text, nothing draws the eye more powerfully than a contrast between light and dark colors. This is why a bold font creates more emphasis than an italic font. (See also bold or italic.)
>
> 2. The perceived intensity of colored type depends not just on the color, but also the size and weight of the font. So a thin or small font can carry a more intense color than a heavy or large font.
>
> 3. I'm not saying it can never be done well, but when someone puts colored type on a colored background, I usually wish they hadn't.
>
> Body text in printed documents (e.g., résumés, research papers, letters) must always be set in black type. No exceptions.
>
> At a typical body-text point size, color isn't effective as a form of emphasis. Small letterforms don't cover much surface area on the page, so colored text isn't noticed unless it's loud.

8.12 Color Blockquotes

owasp_pdf has an extension feature called "color blockquotes." To create a color blockquote, start a line with greater than >. Color blockquotes are immediately followed by four optional parameters separated by a vertical bar |, then a space character. Color blockquotes can not be nested, but two space characters can be used to add one level of indentation (see 8.4 First Line Indent). For color names, see 12. Color Palette.

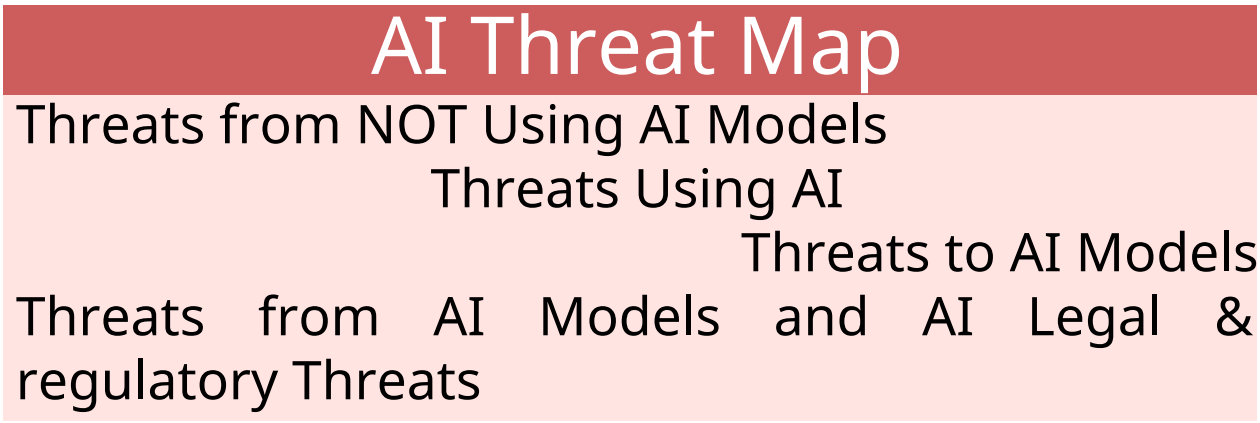
The tag > can be immediately followed by five optional parameters separated by a

vertical bar |:

1. background color name (string),
2. text color name (string),
3. text alighment (string: "left", "center", "right", "justified"),
4. font size (integer in points), and
5. line pitch (integer in points)

Example:

```
>indianred|white|center|30|34 AI Threat Map
>mistyrose|black|left|20|24 Threats from NOT Using AI Models
>mistyrose|black|center|20|24 Threats Using AI Models
>mistyrose|black|right|20|24 Threats to AI Models
>mistyrose|black|justified|20|24 Threats from AI Models and AI Legal & regulatory Threats
```



9. Batch Processing

You can build all registered languages of one project with one command:

```
$ ./owasp_pdf -a <proj code>
```

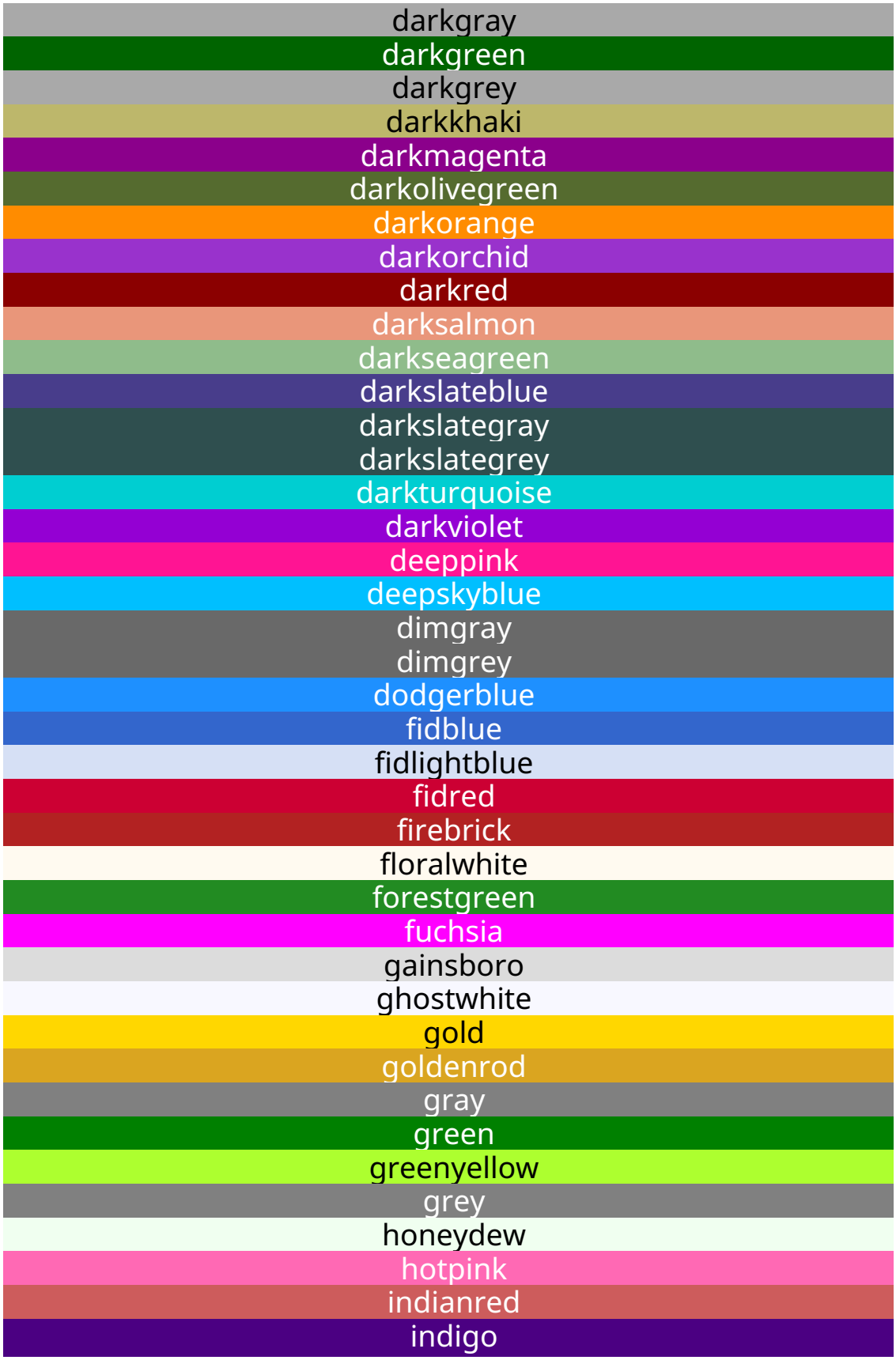
Suppose you have three projects, LLM, GOV, and OLM. You can build all the languages of all the projects in three processes with one command line. macOS Activity Monitor will show three processes of owasp_pdf:

```
$ ./owasp_pdf -a LLM & ./owasp_pdf -a GOV & ./owasp_pdf -a OLM
```

```

## 10. Semantic Versioning

owasp_pdf version number such as v3.7.2 follows the semantic versioning rule.
Given a version number MAJOR.MINOR.PATCH, increment the:

- ▷ MAJOR version when we make incompatible changes to custom pdf json
- ▷ MINOR version when we add functionality which impacts custom pdf json in a backward compatible manner
- ▷ PATCH version when we make backward compatible bug fixes, i.e. no changes to custom pdf json syntax and semantics.

The version number with build id is placed at the lower left corner of every page.

## 11. Language Codes

```

ar-SA : Arabic
be-BY : Belarusian
cs-CZ : Czech
da-DK : Danish
de-DE : German
en-GB : English (United Kingdom)
en-US : English (United States)
en-ZZ : English (International)
es-ES : Spanish (Spain)
es-MX : Spanish (Mexico)
et-EE : Estonian
fi-FI : Finnish
fr-CA : French (Canada)
fr-FR : French (Standard)
he-IL : Hebrew
hi-IN : Hindi
hu-HU : Hungarian
it-IT : Italian
ja-JP : Japanese
ko-KR : Korean
lt-LT : Lithuanian
lv-LV : Latvian
ms-MY : Malaysian
nl-NL : Dutch

no-NO : Norwegian
pl-PL : Polish
pt-BR : Portuguese (Brazil)
pt-PT : Portuguese (Portugal)
ru-RU : Russian
sv-SE : Swedish
th-TH : Thai
tr-TR : Turkish
vi-VN : Vietnamese
zh-CN : Chinese (Simplified)
zh-TW : Chinese (Traditional)
```

## 12. Color Palette

aliceblue
antiquewhite
aqua
aquamarine
azure
beige
bisque
black
blanchedalmond
blue
blueviolet
brown
burlywood
cadetblue
chartreuse
chocolate
coral
cornflower
cornflowerblue
cornsilk
crimson
cyan
darkblue
darkcyan
darkgoldenrod
```


darkgray
darkgreen
darkgrey
darkkhaki
darkmagenta
darkolivegreen
darkorange
darkorchid
darkred
darksalmon
darkseagreen
darkslateblue
darkslategray
darkslategrey
darkturquoise
darkviolet
deeppink
deepskyblue
dimgray
dimgrey
dodgerblue
fidblue
fidlightblue
fidred
firebrick
floralwhite
forestgreen
fuchsia
gainsboro
ghostwhite
gold
goldenrod
gray
green
greenyellow
grey
honeydew
hotpink
indianred
indigo

ivory
khaki
lavender
lavenderblush
lawngreen
lemonchiffon
lightblue
lightcoral
lightcyan
lightgoldenrodyellow
lightgreen
lightgrey
lightpink
lightsalmon
lightseagreen
lightskyblue
lightslategray
lightslategrey
lightsteelblue
lightyellow
lime
limegreen
linen
magenta
maroon
mediumaquamarine
mediumblue
mediumorchid
mediumpurple
mediumseagreen
mediumslateblue
mediumspringgreen
mediumturquoise
mediumvioletred
midnightblue
mintcream
mistyrose
moccasin
navajowhite
navy


oldlace
olive
olivedrab
orange
orangered
orchid
palegoldenrod
palegreen
paleturquoise
palevioletred
papayawhip
peachpuff
peru
pink
plum
powderblue
purple
red
rosybrown
royalblue
saddlebrown
salmon
sandybrown
seagreen
seashell
sienna
silver
skyblue
slateblue
slategray
slategrey
snow
springgreen
steelblue
tan
teal
thistle
tomato
turquoise
violet

wheat
white
whitesmoke
yellow
yellowgreen

13. Appendices

```
APPENDIX 1.  ./owasp_pdf -h  (--help)
================================

usage: owasp_pdf [-h] [-v] [-r] [-s] [-l] [-a]

optional arguments:
-h, --help            show this help message and exit
-v, --version         show owasp_pdf package version
-r, --reg, --registered show registered languages for each project
-s, --silent          disable all print statements
-l                 {LLM_ar-SA,LLM_en-US,LLM_en-ZZ,LLM_es-ES,LLM_fr-FR,LLM_hi-IN,LLM_it-IT,LLM_ja-JP,LLM_pt-BR,LLM_zh-CN,GOV_ar-SA,GOV_be-BY,GOV_cs-CZ,GOV_da-DK,GOV_de-DE,GOV_en-GB,GOV_en-US,GOV_en-ZZ,GOV_es-ES,GOV_es-MX,GOV_et-EE,GOV_fi-FI,GOV_fr-CA,GOV_fr-FR,GOV_he-IL,GOV_hi-IN,GOV_hu-HU,GOV_it-IT,GOV_ja-JP,GOV_ko-KR,GOV_lt-LT,GOV_lv-LV,GOV_ms-MY,GOV_nl-NL,GOV_no-NO,GOV_pl-PL,GOV_pt-BR,GOV_pt-PT,GOV_ru-RU,GOV_sv-SE,GOV_th-TH,GOV_tr-TR,GOV_vi-VN,GOV_zh-CN,GOV_zh-TW,OLM_en-US,OLM_en-ZZ,OLM_ja-JP,LLM_ar-SA,LLM_en-US,LLM_en-ZZ,LLM_es-ES,LLM_fr-FR,LLM_hi-IN,LLM_it-IT,LLM_ja-JP,LLM_pt-BR,LLM_zh-CN,GOV_ar-SA,GOV_be-BY,GOV_cs-CZ,GOV_da-DK,GOV_de-DE,GOV_en-GB,GOV_en-US,GOV_en-ZZ,GOV_es-ES,GOV_es-MX,GOV_et-EE,GOV_fi-FI,GOV_fr-CA,GOV_fr-FR,GOV_he-IL,GOV_hi-IN,GOV_hu-HU,GOV_it-IT,GOV_ja-JP,GOV_ko-KR,GOV_lt-LT,GOV_lv-LV,GOV_ms-MY,GOV_nl-NL,GOV_no-NO,GOV_pl-PL,GOV_pt-BR,GOV_pt-PT,GOV_ru-RU,GOV_sv-SE,GOV_th-TH,GOV_tr-TR,GOV_vi-VN,GOV_zh-CN,GOV_zh-TW,OLM_en-US,OLM_en-ZZ,OLM_ja-JP}, --lang {LLM_ar-SA,LLM_en-US,LLM_en-ZZ,LLM_es-ES,LLM_fr-FR,LLM_hi-IN,LLM_it-IT,LLM_ja-JP,LLM_pt-BR,LLM_zh-CN,GOV_ar-SA,GOV_be-BY,GOV_cs-CZ,GOV_da-DK,GOV_de-DE,GOV_en-GB,GOV_en-US,GOV_en-ZZ,GOV_es-ES,GOV_es-MX,GOV_et-EE,GOV_fi-FI,GOV_fr-CA,GOV_fr-FR,GOV_he-IL,GOV_hi-IN,GOV_hu-HU,GOV_it-IT,GOV_ja-JP,GOV_ko-KR,GOV_lt-LT,GOV_lv-LV,GOV_ms-MY,GOV_nl-NL,GOV_no-NO,GOV_pl-PL,GOV_pt-BR,GOV_pt-PT,GOV_ru-RU,GOV_sv-SE,GOV_th-TH,GOV_tr-TR,GOV_vi-VN,GOV_zh-CN,GOV_zh-TW,OLM_en-US,OLM_en-ZZ,OLM_ja-JP,LLM_ar-SA,LLM_en-US,LLM_en-ZZ,LLM_es-ES,LLM_fr-FR,LLM_hi-IN,LLM_it-IT,LLM_ja-JP,LLM_pt-BR,LLM_zh-CN,GOV_ar-SA,GOV_be-BY,GOV_cs-CZ,GOV_da-DK,GOV_de-
```

```
DE,GOV_en-GB,GOV_en-US,GOV_en-ZZ,GOV_es-ES,GOV_es-MX,GOV_et-EE,GOV_fi-FI,GOV_fr-CA,GOV_fr-FR,GOV_he-IL,GOV_hi-IN,GOV_hu-HU,GOV_it-IT,GOV_ja-JP,GOV_ko-KR,GOV_lt-LT,GOV_lv-LV,GOV_ms-MY,GOV_nl-NL,GOV_no-NO,GOV_pl-PL,GOV_pt-BR,GOV_pt-PT,GOV_ru-RU,GOV_sv-SE,GOV_th-TH,GOV_tr-TR,GOV_vi-VN,GOV_zh-CN,GOV_zh-TW,OLM_en-US,OLM_en-ZZ,OLM_ja-JP}
        set project/language to build PDF for, e.g., GOV_en-ZZ, LLM_ja-JP
-a {LLM,GOV,OLM}, --all {LLM,GOV,OLM}
        process all registered languages for the project at once
```

```
APPENDIX 2.   custom_data_GOV_en-US.json
================================

{
        "doc_cover": true,
        "doc_title": [
                "LLM AI Security &",
                "Governance Checklist"
        ],
        "doc_title_pivot.pt_y": 324.0,
        "doc_subtitles": [
                "",
                "",
                "From the OWASP Top 10",
                "for LLM Applications Team",
                "",
                "",
                "",
                "",
                "",
                "Version: 1.1",
                "Published: March 31, 2024"
        ],
        "doc_header": "",
        "doc_header_pivot.pt_x": 36.0,
        "doc_legal_notice": true,
        "doc_legal_notice_words": [
                "The information provided in this document does not, and is not intended to, constitute legal advice.",
                "All information is for general informational purposes only.",
                "",
```

```
            "This document contains links to other third-party websites. Such links are only for convenience",
            "and OWASP does not recommend or endorse the contents of the third-party sites.",
            "",
            "This project is licensed under the terms of the Creative Commons Attribution-ShareAlike 4.0",
            "International License. ( https://creativecommons.org/licenses/by-sa/4.0/ )"
    ],
    "doc_toc": true,
    "doc_authors_toc": false,
    "doc_toc_contents_title": "Contents",
    "doc_toc_figures_title": "Figures",
    "doc_watermark": true,
    "doc_title_font.size": 40,
    "doc_title_font.line_pitch": 48.0,
    "doc_subtitle_font.size": 20,
    "doc_subtitle_font.line_pitch": 24.0,
    "chapter_pivot.pt_y": 28.1,
    "chapter_font.size": 24,
    "section_font.size": 13,
    "section_font.line_pitch": 18.2,
    "block_font.size": 11,
    "block_font.line_pitch": 15.4,
    "block6_font.size": 11,
    "block6_font.line_pitch": 15.4,
    "caption_font.size": 11,
    "caption_font.line_pitch": 15.4,
    "body_font.size": 11,
    "body_font.line_pitch": 15.4,
    "body_font.line_alignment": "justified",
    "reference_font.size": 13,
    "reference_font.line_pitch": 18.2,
    "reference_font.line_alignment": "justified",
    "blockquote_font.size": 10,
    "blockquote_font.line_pitch": 14.0,
    "blockquote_font.line_alignment": "justified",
    "unordered_list_marker": "square",
    "md_file_range": [
            0,
            999999
```

```
            ],
            "max_image_scale": 2.0
}
```