



W9 Labeling Instructions

Evaluation Set Requirements



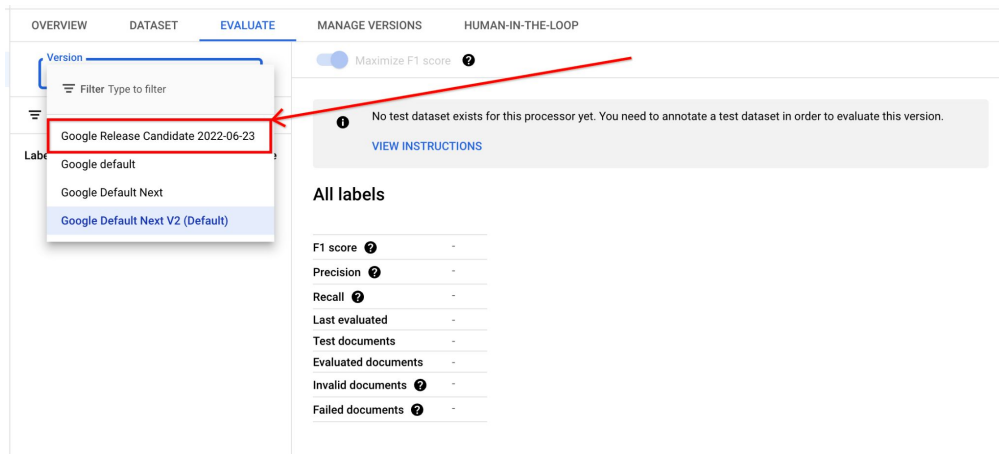
In order to do fair and thorough evaluation of the model, the evaluation set should take the following dimensions into consideration

1. **Revisions:** Documents should be in English for both (Rev. 10-2018, Rev. 11-2017) revisions.
2. **Size:** It is recommended to have 500+ documents for each evaluation set.
3. **Template Diversity:** The documents should be diverse, e.g. different templates, issue year, different sources, states (if applicable) ... etc.
4. **Value Diversity:** Documents should capture all possible values. E.g. documents should cover all checkbox options for “Federal Tax Classification”. Same applies to all ENUM or optional entities.
5. **Handwritten:** It is preferred to separate handwritten documents from typed ones (i.e. have 2 separate evaluation sets) and report metrics for each.

Processor Version

These instructions are specific for W9 Processor version:

“Google Release Candidate 2022-06-23”



The screenshot displays the 'EVALUATE' tab of the W9 Processor interface. The top navigation bar includes 'OVERVIEW', 'DATASET', 'EVALUATE', 'MANAGE VERSIONS', and 'HUMAN-IN-THE-LOOP'. A 'Version' dropdown menu is open, showing a list of processor versions: 'Google Release Candidate 2022-06-23', 'Google default', 'Google Default Next', and 'Google Default Next V2 (Default)'. A red box highlights the 'Google Release Candidate 2022-06-23' option, and a red arrow points from this option to a message box on the right. The message box contains the text: 'No test dataset exists for this processor yet. You need to annotate a test dataset in order to evaluate this version.' Below the message box is a 'VIEW INSTRUCTIONS' link. The 'All labels' section is visible, showing a table with metrics: F1 score, Precision, Recall, Last evaluated, Test documents, Evaluated documents, Invalid documents, and Failed documents, all with a '-' value.

Metric	Value
F1 score	-
Precision	-
Recall	-
Last evaluated	-
Test documents	-
Evaluated documents	-
Invalid documents	-
Failed documents	-

Entity List

Field	Description	Base Type
name		string
business_name		string
address		string
city_state_zip		string
account_numbers		string
ltc_tax_classification		string
federal_tax_classification_other	Text field that is usually populated when "Other" checkbox option is selected.	string
exempt_payee_code		string
exempt_fatca_code		string
ssn	Value should be composed of digits only.	number
ein	Value should be composed of digits only.	number
has_signature	Enum: YES, NO (Case Sensitive)	string (enum)
has_signature_date	Enum: YES, NO (Case Sensitive)	string (enum)
form_revision_date		string
federal_tax_classification	Checkbox with 7 options one of them should be selected. Enum (Case Sensitive): IndividualsoleproprietororsinglememberLLC, CCorporation, SCorporation, Partnership, Trustestate, Limitedliabilitycompany, Other	string (enum)

Note: There will be zero or one instance of each entity.

form_revision_date

name

business_name

federal_tax_classification = "Other"

address

city_state_zip

account_numbers

has_signature = "NO"



Request for Taxpayer Identification Number and Certification

Give Form to the requester. Do not send to the IRS.

Go to www.irs.gov/FormW9 for instructions and the latest information.

1 Name (as shown on your income tax return). Name is required on this line; do not leave this line blank.
John Doe

2 Business name (disregarded entity name, if different from above)
Fake Business Name

3 Check appropriate box for federal tax classification of the person whose name is entered on line 1. Check only one of the following seven boxes.

Individual/sole proprietor or single-member LLC C Corporation S Corporation Partnership Trust/estate

Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=Partnership) **Other**

Note: Check the appropriate box in the line above for the tax classification of the single-member owner. Do not check LLC if the LLC is classified as a single-member LLC that is disregarded from the owner unless the owner of the LLC is another LLC that is not disregarded from the owner for U.S. federal tax purposes. Otherwise, a single-member LLC that is disregarded from the owner should check the appropriate box for the tax classification of its owner.

4 Exemptions (codes apply only to certain entities, not individuals; see instructions on page 3):
 Exempt payee code (if any) **A**
 Exemption from FATCA reporting code (if any) **B**
 (Applies to accounts maintained outside the U.S.)

5 Address (number, street, and apt. or suite no.) See instructions.
123 Fake St.

6 City, state, and ZIP code
New York, NY, 12345

7 List account number(s) here (optional)
12345, 67890

exempt_payee_code

ltc_tax_classification

exempt_fatca_code

federal_tax_classification_other

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on line 1 to avoid backup withholding. For individuals, this is generally your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the instructions for Part I, later. For other entities, it is your employer identification number (EIN). If you do not have a number, see How to get a TIN, later.

Social security number

1	2	3	-	4	5	-	6	7	8	9
---	---	---	---	---	---	---	---	---	---	---

SSN

or

Employer identification number

1	2	-	3	4	5	6	7	8	9	
---	---	---	---	---	---	---	---	---	---	--

ein

Note: Either SSN or EIN should be present

Part II Certification

Under penalties of perjury, I certify that:

- The number shown on this form is my correct taxpayer identification number (or I am waiting for a number to be issued to me); and
- I am not subject to backup withholding because: (a) I am exempt from backup withholding, or (b) I have not been notified by the Internal Revenue Service (IRS) that I am subject to backup withholding as a result of a failure to report all interest or dividends, or (c) the IRS has notified me that I am no longer subject to backup withholding; and
- I am a U.S. citizen or other U.S. person (defined below); and
- The FATCA code(s) entered on this form (if any) indicating that I am exempt from FATCA reporting is correct.

Certification instructions. You must cross out item 2 above if you have been notified by the IRS that you are currently subject to backup withholding because you have failed to report all interest and dividends on your tax return. For real estate transactions, item 2 does not apply. For mortgage interest paid, acquisition or abandonment of secured property, cancellation of debt, contributions to an individual retirement arrangement (IRA), and generally, payments other than interest and dividends, you are not required to sign the certification, but you must provide your correct TIN. See the instructions for Part II, later.

Sign Here Signature of U.S. person **[Signature]** Date **[Date]**

Has_signature_date = "NO"

General Instructions

Section references are to the Internal Revenue Code unless otherwise noted.

Future developments. For the latest information about developments related to Form W-9 and its instructions, such as legislation enacted after they were published, go to www.irs.gov/FormW9.

Purpose of Form

An individual or entity (Form W-9 requester) who is required to file an information return with the IRS must obtain your correct taxpayer identification number (TIN) which may be your social security number (SSN), individual taxpayer identification number (ITIN), adoption taxpayer identification number (ATIN), or employer identification number (EIN), to report on an information return the amount paid to you, or other amount reportable on an information return. Examples of information returns include, but are not limited to, the following.

- Form 1099-DIV (dividends, including those from stocks or mutual funds)
- Form 1099-MISC (various types of income, prizes, awards, or gross proceeds)
- Form 1099-B (stock or mutual fund sales and certain other transactions by brokers)
- Form 1099-S (proceeds from real estate transactions)
- Form 1099-K (merchant card and third party network transactions)
- Form 1098 (home mortgage interest), 1098-E (student loan interest), 1098-T (tuition)
- Form 1099-C (canceled debt)
- Form 1099-A (acquisition or abandonment of secured property)

Use Form W-9 only if you are a U.S. person (including a resident alien), to provide your correct TIN.

If you do not return Form W-9 to the requester with a TIN, you might be subject to backup withholding. See What is backup withholding, later.



General Labeling Instructions

Occurrence Type



There are 4 occurrence types which restricts the number of annotations for each entity:

- **OPTIONAL_ONCE**: Entity can be present zero or multiple times but should have one unique value.
- **OPTIONAL_MULTIPLE**: Entity will appear zero or multiple times and with different values.
- **REQUIRED_ONCE**: Entity can be present once or multiple times but should have one unique value.
- **REQUIRED_MULTIPLE**: Entity will appear once or multiple times and with different values.

All W9 form entities belongs to any of the occurrence type mentioned above .

Optical Character Recognition (OCR) Errors



Drawing bounding boxes around entities will automatically populate a text captured through OCR (Optical Character Recognition). In some cases, the captured text is not accurate and doesn't reflect what is on the page, especially for documents that are blurry or of low resolution. These values should be corrected manually through "Value" input text.

Bounding Box Annotation

Bounding boxes should fit the correct text as much as possible:

- Avoid overlapping bounding boxes. This can be tricky when values are close to each other.
- Avoid oversized boxes which incorporates background artifacts such as table borders.

Example 1

See Specific: is disregarded from the owner should check the appropriate

Other (see instructions) ▶ Hernandez Junction

5 Address (number, street, and apt. or suite no.) See instruction
1983 State Street

6 City, state, and ZIP code
Freeport, NY 11520

7 List account number(s) here (optional)
355786KYL

Part I Taxpayer Identification Number (TIN)

Incorrect Annotation:
Bounding boxes are overlapping

Example 2

See Specific: is disregarded from the owner should check the appropriate

Other (see instructions) ▶ Hernandez Junction

5 Address (number, street, and apt. or suite no.) See instruction
1983 State Street

6 City, state, and ZIP code
Freeport, NY 11520

7 List account number(s) here (optional)
355786KYL

Part I Taxpayer Identification Number (TIN)

Correct Annotation

Example 3

See Specific: is disregarded from the owner should check the appropriate

Other (see instructions) ▶ Hernandez Junction

5 Address (number, street, and apt. or suite no.) See instruction
1983 State Street

6 City, state, and ZIP code
Freeport, NY 11520

7 List account number(s) here (optional)
355786KYL

Part I Taxpayer Identification Number (TIN)

Incorrect Annotation:
Bounding boxes oversized

Case Sensitivity



DocAI evaluation is case sensitive and can affect the final evaluation. While annotation please be aware to fix any casing issues generated by OCR.

Please note that ENUM values are also case sensitive and the values should be fixed as listed in the Entity List slide.

Examples:

- "John doe" != "John Doe"
- "Ccorporation" != "CCorporation"

Numbers

Numbers and Account Numbers such as SSN and EIN should only be composed of digits. In the example below, the OCR captured extra dashes and spaces that should be manually removed.

The image shows a PDF viewer interface on the left and a scanned W-9 form on the right. The PDF viewer has a search bar with the text 'synthetic_file_1655375338027891.pdf'. Below the search bar is a filter section with a 'Filter' button and a text input field. The 'Label *' field contains 'ssn' and the 'Value' field contains '944-28 - 6745'. Below these fields are three buttons: 'CANCEL', 'DELETE', and 'CONFIRM'. A red arrow points from the 'Value' field to the 'ssn' label in the form below.

The W-9 form is titled 'Request for Taxpayer Identification Number and Certification'. It includes the following information:

- Form W-9 (Rev. October 2018) Department of the Treasury Internal Revenue Service
- 1 Name (as shown on your income tax return). Name is required on this line; do not leave this line blank. Grant, O.
- 2 Business name/disregarded entity name, if different from above TYLER LTD
- 3 Check appropriate box for federal tax classification of the person whose name is entered on line 1. Check only one of the following seven boxes:
 - Individual/sole proprietor or single-member LLC
 - C Corporation
 - S Corporation
 - Partnership
 - Trust/estate
 - Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=Partnership) **S**
 - Other (see instructions) **Danny Key**
- 4 Exemptions (codes apply only to certain entities, not individuals; see instructions on page 3):
 - Exempt payee code (if any) **6**
 - Exemption from FATCA reporting code (if any) **A**
- 5 Address (number, street, and apt. or suite no.) (See instructions.) 1568 Prospect Valley Road
- 6 City, state, and ZIP code Round Lake, IL 60073
- 7 List account number(s) here (optional) 464741GHT

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on line 1 to avoid backup withholding. For individuals, this is generally your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the instructions for Part I, later. For other entities, it is your employer identification number (EIN). If you do not have a number, see *How to get a TIN*, later.

Social security number

9 4 4 - 2 8 - 6 7 4 5

Money

For money entries the annotation bounding box and value text should include the currency if present on the document.

If the document contains currency with no values, then annotations could be omitted, as shown in the example on the bottom right.

Contract Summary

	This Quarter	Year To Date	Since Issue Date ¹
Beginning Value on 09/28/2018	\$201,931.64		
Total Premium	\$0.00	\$4,000.00	\$146,388.89
Total Withdrawals	\$0.00	\$0.00	\$0.00
Total Tax Withheld	\$0.00	\$0.00	
Net Change	(\$23,104.00)	(\$15,683.48)	
Ending Value ² on 12/31/2018	\$178,827.64		

FATCA filing requirement <input type="checkbox"/>	12 State tax withheld \$	13 State/Payer's state no. 000	14 State distribution \$
Date of Payment	15 Local tax withheld \$	16 Name of locality	17 Local distribution \$
	\$		\$

www.irs.gov/form 1099r

Department of the Treasury - Internal Revenue Service

Empty Values

In general, if the value does not exist in the form then there is no need to label it. In the example below `ltc_tax_classification` should not be annotated since there is no value present in the designated field.

It is also possible to annotate empty entities (with empty values) but that will not affect the final metrics.

Form W-9 (Rev. October 2018) Department of the Treasury Internal Revenue Service		Request for Taxpayer Identification Number and Certification		Give Form to the requester. Do not send to the IRS.	
1 Name (as shown on your income tax return). Name is required on this line; do not leave this line blank. John Doe					
2 Business name/disregarded entity name, if different from above Fake Business Name					
Print or type. See specific instructions on page 3.	3 Check appropriate box for federal tax classification of the person whose name is entered on line 1. Check only one of the following seven boxes.			4 Exemptions (codes apply only to certain entities, not individuals; see instructions on page 3).	
	<input type="checkbox"/> Individual/sole proprietor or single-member LLC			Exempt payer code (if any) <u>A</u>	
	<input type="checkbox"/> Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=Partnership) ▶ _____			Exemption from FATCA reporting code (if any) <u>B</u>	
	<input checked="" type="checkbox"/> Other (see instructions) ▶ _____ Other Type _____			(Applies to accounts maintained outside the U.S.)	
5 Address (number, street, and apt. or suite no.) See instructions. 123 Fake St.				Requester's name and address (optional)	
6 City, state, and ZIP code New York, NY, 12345					
7 List account number(s) here (optional)				12345, 67890	
Taxpayer Identification Number (TIN)					

`ltc_tax_classification`

Note: There is one exception to this rule, and will be explained in the next slide.

YES/NO Values

When annotating ENUM types that includes YES/NO options such as `has_signature`, and `has_signature_date`, a bounding box must be drawn at that field and the value should be changed to either “YES” if the value exists or “NO” if it does not (**Note: these values are case sensitive**). This will allow the groundtruth values to match the processor predicted output. Examples:

Example 1: YES

The screenshot shows a PDF form titled "Request for Taxpayer Identification Number and Certification" (W-9). The form is for Jose Sarah V. The "Value" field in the left sidebar is highlighted with a blue bounding box and contains the text "YES". A red arrow points from the "Value" field to the "Taxpayer Identification Number (TIN)" section of the form, which contains the number 83754220V. The form also includes sections for "Certification" and "General Instructions".

Example 2: NO

The screenshot shows a PDF form titled "Request for Taxpayer Identification Number and Certification" (W-9). The form is for Elizabeth Samuel David. The "Value" field in the left sidebar is highlighted with a blue bounding box and contains the text "NO". A red arrow points from the "Value" field to the "Taxpayer Identification Number (TIN)" section of the form, which contains the number 318-75-2481. The form also includes sections for "Certification" and "General Instructions".

Checkbox Values For Evaluation On Pretrained Versions

Checkbox entities are tricky to annotate. The annotation required for evaluation is different than the annotation for training.

For Evaluation: Checkbox annotation needs to be drawn around the option that is *selected*. Then the *Value* should be changed to one of the ENUM options listed in the Field Entity List slide (the value is case sensitive)

The image shows a PDF viewer displaying Form W-9, 'Request for Taxpayer Identification Number and Certification'. The form is for 'Grant, O.' and 'TYLER LTD'. In the 'Value' field, the text 'Other' is entered, and a red arrow points to it from the left. The 'CONFIRM' button is highlighted in blue.

Example 1: Other option is selected

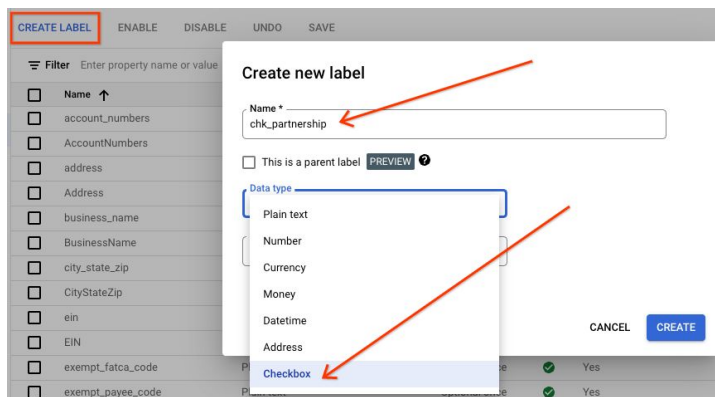
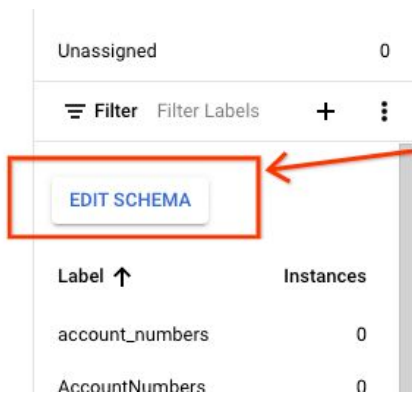
The image shows a PDF viewer displaying Form W-9, 'Request for Taxpayer Identification Number and Certification'. The form is for 'Deborah Melissa-Sylvia' and 'Valencia Llc'. In the 'Value' field, the text 'Trust/estate' is entered, and a blue box highlights it. The 'CONFIRM' button is highlighted in blue.

Example 2: Trust/estate option is selected

Checkbox Values For Training & Eval on uptrain - 1/2

For training, checkbox entities should be added to the dataset schema first by following the steps:

1. Click on “Edit Schema” button.
2. Click on “Create Label”
3. Type a suitable name, change “Data Type” to “Checkbox” and click “Create”.
4. Repeat for all checkbox options.



Checkbox Values For Training & Eval on uptrain - 2/2

During annotation for training, all checkbox options should be labeled with the correct value as shown in the example below:

- chk_c_corp
- chk_partnership
- chk_s_corp
- chk_trust_estate
- chk_sole_proprietor
- chk_llc
- chk_other_form

Internal Revenue Service ▶ Go to www.irs.gov/FormW9 for instructions and the latest information.

1 Name (as shown on your income tax return). Name is required on this line; do not leave this line blank.
Donna Erin

2 Business name/disregarded entity name, if different from above
Mcdowell-ross

3 Check appropriate box for federal tax classification of the person whose name is entered on line 1. Check only **one** of the following seven boxes.

Individual/sole proprietor or single-member LLC C Corporation S Corporation Partnership Trust/estate

Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=Partnership) ▶ **P**

Note: Check the appropriate box in the line above for the tax classification of the single-member owner. Do not check LLC if the LLC is classified as a single-member LLC that is disregarded from the owner unless the owner of the LLC is another LLC that is **not** disregarded from the owner for U.S. federal tax purposes. Otherwise, a single-member LLC that is disregarded from the owner should check the appropriate box for the tax classification of its owner.

Other (see instructions) ▶ **Tyler Station**

5 Address (number, street, and apt. or suite no.) See instructions. Requester's name an
2549 Rainy Day Drive

6 City, state, and ZIP code
Round Lake, IL 60073

Print or type. See Specific Instructions on page 3.